

OM of: US-09-674-779-2 to: GenEmbl:* out_format : pfs

Date: Sep 16, 2002 7:55 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame.p2n.pool-DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09674779/runat_12092002_124143_26862/app_query.fasta_1.310
-DB=GenEmbl -QFMT=fastcap -SUFFIX=oli.rge -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -GAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -CGN1_1_4034 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-2

Query length: 250

Database: GenEmbl.*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 1844.010000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGNAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
gb_pat:AX081145	+	250.00	5248.11	5.8e-284	753	AX081145 Sequence 1 from Patent
gb_pat:AX067442	+	250.00	5221.80	1.7e-282	23210	AX067442 Sequence 17 from Patent
gb_htg:AC097141	-	9.00	120.01	254.01	157901	AC097141 Rattus norvegicus c
gb_pr:AC068544	-	9.00	119.94	256.38	159397	AC068544 Homo sapiens BAC cl
gb_htg:AC106164	-	9.00	119.05	287.28	178928	AC106164 Rattus norvegicus c
gb_htg:AC026821	+	9.00	118.74	299.06	186385	AC026821 Homo sapiens chromo
gb_pr:AL335499	-	9.00	118.27	317.62	198141	AL335499 Human DNA sequence
gb_htg:AL512373	-	9.00	117.83	335.99	209790	AL512373 Homo sapiens chromo
gb_pat:AX081147	+	8.00	162.46	1.10	40	AX081147 Sequence 3 from Patent
gb_pat:AX081148	+	8.00	139.35	1.64	60	AX081148 Sequence 4 from Patent
gb_sts:G22057	-	8.00	144.23	11.37	430	G22057 human STS WI-13488, sequ
gb_sts:G61482	-	8.00	141.06	17.07	650	G61482 SHGC-85859 Human Homo sa
em_htgo_inv:AC085197	+	8.00	138.96	22.33	854	AC085197 Giardia intestinalis c
gb_ba:MSGLIPOPRE	+	8.00	138.56	23.52	900	L08229 Mycobacterium intracellu
em_htgo_inv:AC028827	+	8.00	137.74	24.52	939	AC028827 Giardia intestinalis c
em_htgo_inv:AC071017	+	8.00	137.96	25.42	974	AC071017 Giardia intestinalis c
em_htgo_inv:AC060939	+	8.00	137.88	25.68	984	AC060939 Giardia intestinalis c
em_htgo_inv:AC076436	+	8.00	137.74	26.14	1002	AC076436 Giardia intestinalis c
gb_pl:ATH271473	-	8.00	136.19	31.88	1226	AT271473 Arabidopsis thaliana
gb_pat:E03308	-	8.00	135.97	32.80	1262	E03308 DNA encoding recombinan
gb_pl:AF370181	-	8.00	135.13	36.51	1407	AF370181 Arabidopsis thaliana
gb_ov:JPJ00267	+	8.00	132.94	48.36	1872	AFJ00267 Icterus punctatus m
gb_pat:AR156086	+	8.00	132.41	51.79	2007	AR156086 Sequence 7 from patent
gb_vl:FCCPKS19	-	8.00	132.39	51.87	2010	X99446 Feline calicivirus gene
gb_vl:ATH271473	-	8.00	132.30	52.48	2034	E12464 Capsid protein gene of
gb_pat:E12464	+	8.00	132.14	53.57	2077	AK022355 Homo sapiens mRNA for
gb_pr:AC022355	+	8.00	131.55	57.78	2243	AB010463 Vibrio parahaemolytic
gb_ba:AB010463	+	8.00	129.51	75.09	2927	M32819 Feline calicivirus caps
gb_vl:CVXCAPS	+	8.00	128.14	89.54	3500	AF072657 Danio rerio HLB-conta
gb_ov:AF072657	+	8.00	128.10	89.95	3516	AF032952 Homo sapiens nCAP-lik
gb_vl:FCLJ3E	-	8.00	127.53	96.84	3790	D90357 Feline calicivirus (FCV
gb_vl:AF051530	-	8.00	123.84	155.42	6128	AB051530 Homo sapiens mRNA for
gb_ro:AF236130	-	8.00	123.64	159.48	6291	AF236130 Rattus norvegicus DOC
gb_pat:AX251205	+	8.00	122.11	194.02	6777	AX251205 Sequence 173 from Pat
gb_vl:FCU13992	-	8.00	122.11	194.02	6777	U13992 Feline calicivirus CF1
gb_vl:FCLF4	-	8.00	121.45	211.17	8367	D31836 Feline calicivirus gene
gb_ba:AF400582	+	8.00	121.45	211.17	8367	AF400582 Actinobacter sp. ADP
gb_ba:AE002287	+	8.00	119.50	271.00	10780	AE002287 Chlamydia muridarum

gb_ba:AE003962 - 8.00 119.36 276.10 10986 ! AE003962 Xylella fastidios
gb_ba:AE001354 + 8.00 118.83 295.44 11768 ! AE001354 Chlamydia trachom
gb_htg:AC110326 - 8.00 115.28 466.00 18696 ! AC110326 Rattus norvegicus
gb_in:CEK07A1 - 8.00 109.62 963.09 39086 ! Z81097 Caenorhabditis eleg
gb_pr:AL357372 - 8.00 108.79 1.1e+03 43553 ! AL357372 Human DNA sequenc
gb_htg:AC098224 - 8.00 108.63 1.1e+03 44428 ! AC098224 Rattus norvegicus
gb_htg:AC014969 - 8.00 108.32 1.1e+03 46275 ! AC014969 Drosophila melano

seq_name: gb_pat:AX081145

seq_documentation_block:

LOCUS AX081145 753 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 1 from Patent WO0109335.
ACCESSION AX081145
VERSION AX081145.1 GI:13170040

KEYWORDS Moraxella catarrhalis.

SOURCE Moraxella catarrhalis.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 753)

AUTHORS Thonnard, J.

TITLE Moraxella catarrhalis antigen bas120

JOURNAL Patent: WO 0109335-A 1 08-FEB-2001;

SmithKline Beecham Biologicals s.a. (BE)

FEATURES Location/Qualifiers

source

1..753

/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

BASE COUNT 239 a 166 c 152 g 196 t

ORIGIN

alignment_scores:

Quality: 250.00 Length: 250

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AX081145

Align seg 1/1 to: AX081145 from: 1 to: 753

1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeu 17

|||||

1 ATGAAATTTTATCAATCACTTTAATCACTTATCAGCAGTAGTGTCT 50

|||||

17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34

|||||

51 GGTTCATGTCAGTCACCCATACCAACCAATCCACCAAGTATCCCCAATA 100

|||||

34 yThrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThr 50

|||||

101 AAAGCCATCGGTACTGATTAATAAGATAAATCGGTTCGAGGCATTTTGA 150

|||||

51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGl 67

|||||

151 CATGACGACGATGAATCTGTACCCATGCGGTTCGAGGCATTTTGA 200

|||||

67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84

|||||

201 GACTTGGCTACAGATGCACCATGCCACCAACCAAGAGGTAGTAGGTATC 250

|||||

84 InAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100

|||||

251 AGCGGTATTTACATCAAGACTTGGTAAATATCTGCCACCAATGAGTCAA 300

|||||

101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGl 117

|||||

301 CTACTAACGACTGCACGGTCATGGCAGGATGTGGTCATGAACCTTATCA 350

|||||

117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134

|||||

351 GGTGCCACAGAACATCTTTGGGTCAGATTGTACCAACATGTCACATTGT 400
 134 yTGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
 11583 AAACGGCATCGGTACTGATTAATAAGATAAAATCGGTATCATATACA 11632
 401 ATCAGATCTCAAAAGTAGGGCATATTGCCAGCAATATACCAAAATTCGC 450
 151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
 451 TCAGTTTATCGCAATCTGAACTCAACCAATGTGCTGGTGGTGGAGCTAT 500
 167 tSerLysHisLeuThrAsnSerAlaIleAspTleTyrValProAspLeuG 184
 501 GAGTAGAGCATTTGACCAATAGTGGCTTGTATTTGGGTGGCTGACCTTG 550
 184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
 551 AAATAAAAGCCAGGCACCTGTATGAGTCTCAAAACCGCTATGCCAATAT 600
 201 TrpLeuGluHisGlyGlnAsnGlnAsnPheGlyLeuTyrAlaTh 217
 601 TGGCTAGAGCATGGGGCAAAACCAAAATTTGGGCTGGGTTTATACGCCAC 650
 217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTyrGlyAlaGlnP 234
 651 AGGGGCGATTCTCGATACCAAGGTTTAGAAAATGGGGTGTCTCAAT 700
 234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
 701 TTTCTGAAACAACTCTATTGTCGTCTATGCTTACCAAAAAATAAGCTA 750

seq_name: gb_pat:AX067442

seq_documentation_block:
 LOCUS AX067442 23210 bp DNA linear PAT 24-JAN-2001

DEFINITION Sequence 17 from Patent WO0078968.

ACCESSION AX067442

VERSION AX067442.1 GI:12545062

KEYWORDS Moraxella catarrhalis.

SOURCE Moraxella catarrhalis.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 23210)

AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.

TITLE Nucleotide sequences of moraxella catarrhalis genome

JOURNAL Patent: WO 0078968-A 17 28-DEC-2000;

INCYTE Genomics, Inc. (US)

FEATURES Location/Qualifiers

1..23210

/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

28

/note="a or g or c or t, unknown, or other"

BASE COUNT 7108 a 4482 c 4951 g 6668 t 1 others

ORIGIN

alignment_scores:

Quality: 250.00 Length: 250

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AX067442 ..

Align seg 1/1 to: AX067442 from: 1 to: 23210

1 MetLysAsnPheAsnGlnTyrPheIleThrLeuIleSerSerMetLe 17

11483 ATGAAAATTTTATCACTACTTTATACATACATTTATCAGCAGTATGCT 11532

17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

11533 GGTGTCATGCAGTGCACCATACCAACCAATCCCAAGTATCCCAATAA 11582
 34 yThrProSerValLeuIleThrLysAspLysIleGlyAspHisThr 50
 11583 AAACGGCATCGGTACTGATTAATAAGATAAAATCGGTATCATATACA 11632
 51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
 11633 CATGAGCAGCATGAATCTGTAGCCATGTCGGTTTGCAGGCACATTTTGA 11682
 67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
 11683 GACTTGGCTACAGATGCACCATGCCACCAACAAAGAGGTAGTTAGGTATC 11732
 84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGln 100
 11733 AGGCGTATTTACAAATCAAGACTTGGTAATTAATCTGCCACCAATGAGTCAA 11782
 101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG1 117
 11783 CTACTACGACTGCACGGTTCATGGCAGGCATGTGGTCATGAACCTTATCA 11832
 117 nLeuProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
 11833 GCTGCCACCAAGACATCTTTGGGGTCAGATTTGCCAGCAATATCCCAATTCGC 11882
 134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
 11883 ATCAAGATCTCAAAAGTAGGGCATATTGCCAGCAATATCCCAATTCGC 11932
 151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
 11933 TCAGTTTATCGCAATCTGAACTCAACCAATGTGCTGGTGGTGCAGCTAT 11982
 167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
 11983 GAGTAGAGCATTTGACCAATAGTGGCATTTATTTGGGTGGCTGACCTTG 12032
 184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
 12033 AAATAAAAGCCAGGCACCTGTATGAGCTTCAAAACCGCTATGCCAATAT 12082
 201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
 12083 TGGCTAGAGCATGGCGAAACCAAAATTTGGGTGGTGTATACGCCAC 12132
 217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTyrGlyAlaGlnP 234
 12133 AGGGGCGATTCTCGATACCAAGGTTTAGAAAATGGGGTGTCTCAAT 12182
 234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
 12183 TTTCTGAAACAACTCTATTGTCGTCTATGCTTACCAAAAAATAAGCTA 12232

seq_name: gb_htg:AC097141

seq_documentation_block:

LOCUS AC097141 157901 bp DNA linear HTG 20-DEC-2001

DEFINITION Rattus norvegicus clone CH230-181K9, *** SEQUENCING IN PROGRESS

ACCESSION AC097141

VERSION AC097141.3 GI:17973578

KEYWORDS HTG; HTGS_PHASE1.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 157901)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

1

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lied, C., Liu, J., Liu, W., Loulseghe, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, K., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sison, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, K., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, L., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 157901)
Worley, K.C.

Submitted (11-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064462.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GICJ
Center clone name: CH230-181K9
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to findPhrapList

Consensus quality: 122789 bases at least Q40
Consensus quality: 133264 bases at least Q30
Consensus quality: 142961 bases at least Q20
Estimated insert size: 131949; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 63 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence. as soon as it is available and the accession number will be preserved.

1 4480: contig of 4480 bp in length

4481 4580: gap of unknown length
4581 9598: contig of 5018 bp in length
9599 9698: gap of unknown length
9699 13571: contig of 5873 bp in length
13572 15671: gap of unknown length
15672 20138: contig of 4467 bp in length
20139 20238: gap of unknown length
20239 24260: contig of 4022 bp in length
24261 24360: gap of unknown length
24361 28149: contig of 3789 bp in length
28150 32736: gap of unknown length
32737 32836: gap of unknown length
32837 35463: contig of 2627 bp in length
35464 40210: contig of 4647 bp in length
40211 40310: gap of unknown length
40311 43598: contig of 3288 bp in length
43599 45945: contig of 2247 bp in length
45946 46045: gap of unknown length
46046 49825: contig of 3780 bp in length
49826 54719: contig of 4794 bp in length
54720 54819: gap of unknown length
54820 58816: gap of unknown length
58817 63070: contig of 4254 bp in length
63071 66484: contig of 3314 bp in length
66485 66584: gap of unknown length
66585 69477: contig of 2893 bp in length
69478 71444: contig of 1867 bp in length
71445 71545: gap of unknown length
71546 74171: contig of 2627 bp in length
74172 74272: gap of unknown length
74273 76780: contig of 2509 bp in length
76781 76880: gap of unknown length
76881 79690: contig of 2810 bp in length
79691 82447: contig of 2657 bp in length
82448 82548: gap of unknown length
82549 84890: contig of 2343 bp in length
84891 88153: contig of 3163 bp in length
88154 88253: gap of unknown length
88254 90755: contig of 2502 bp in length
90756 90856: gap of unknown length
90857 93300: contig of 2444 bp in length
93301 96121: contig of 2722 bp in length
96122 96222: gap of unknown length
96223 98033: contig of 1812 bp in length
98034 100571: contig of 2438 bp in length
100572 103441: gap of unknown length
103442 103541: contig of 2770 bp in length
103542 10524: contig of 1983 bp in length
10525 105624: gap of unknown length
105625 107507: contig of 1883 bp in length
107508 109471: contig of 1864 bp in length
109472 109571: gap of unknown length
109572 111494: contig of 1923 bp in length
111495 113470: gap of unknown length
113471 113570: contig of 1876 bp in length
113571 116166: gap of unknown length
116167 116266: gap of unknown length
116267 118544: contig of 2278 bp in length
118545 118644: gap of unknown length

118645 120674: contig of 2030 bp in length
120675 120774: gap of unknown length
120775 122855: contig of 1511 bp in length
122856 122386: gap of unknown length
122387 123737: contig of 1352 bp in length
123738 123837: gap of unknown length
123838 125540: contig of 1703 bp in length
125541 125640: gap of unknown length
125641 126909: contig of 1269 bp in length
126910 127010: gap of unknown length
127011 128225: contig of 1216 bp in length
128226 128326: gap of unknown length
128327 129438: contig of 1113 bp in length
129439 129538: gap of unknown length
129539 131623: contig of 2085 bp in length
131624 131723: gap of unknown length
131724 133593: contig of 1870 bp in length
133594 133693: gap of unknown length
133694 135261: contig of 1568 bp in length
135262 135361: gap of unknown length
135362 136751: contig of 1390 bp in length
136752 136851: gap of unknown length
136852 137975: contig of 1124 bp in length
137976 138075: gap of unknown length
138076 139657: contig of 1582 bp in length
139658 139757: gap of unknown length
139758 140811: contig of 1054 bp in length
140812 140911: gap of unknown length
140912 142409: contig of 1498 bp in length
142410 142509: gap of unknown length
142510 144063: contig of 1554 bp in length
144064 144163: gap of unknown length
144164 145707: contig of 1544 bp in length
145708 145807: gap of unknown length
145808 146876: contig of 1069 bp in length
146877 146976: gap of unknown length
146978 148337: contig of 1361 bp in length
148338 148437: gap of unknown length

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AC097141/rev ..
Align seg 1/1 to reverse of: AC097141 from: 1 to: 157901

6 GlnTyrPheIleThrLeuIleSer 14
52516 CAGTATTATTAACGACCTGTATCT 52490

seq_name: gb_pr:AC068544

seq_documentation_block:
LOCUS AC068544 159397 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-369J9 from 2, complete sequence.
ACCESSION AC068544
VERSION AC068544.7 GI:15920133
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159397)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 159397)
AUTHORS Harris, A., Cordum, H., Dignan, G. and Cotton, M.
TITLE The sequence of Homo sapiens BAC clone RP11-369J9

JOURNAL REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Unpublished (2001)
3 (bases 1 to 159397)
Waterston, R.H.
Direct Submission
Submitted (03-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 159397)
Waterston, R.H.
Direct Submission
Submitted (04-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 159397)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 4, 2001 this sequence version replaced gi:14333968.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics

Center project name: H_NH0369J09

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-308G19; the clone sequenced
to the right is RP11-84A12. Actual start of this clone is at base
position 1 of RP11-369J9; actual end is at base position 159397 of
RP11-369J9.

Data from AC012486 was used to finish this clone, AC068544.
Polymorphisms have been identified between AC068544 and AC012486.

Unresolved dinucleotide repeat from base position 14017 to 14486.
Location/Qualifiers
1. .159397
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

```

/clone="RP11-369J9"
/clone_lib="RPCI-11"
112. 227
/rpt_family="L2"
289. 380
/rpt_family="CR1"
503. 523
/rpt_family="AT_rich"
570. 671
/rpt_family="CR1"
1108. 1263
/rpt_family="L2"
1584. 1648
/rpt_family="GA-rich"
3550. 3605
/rpt_family="AT_rich"
3740. 4235
/rpt_family="L2"
4223. 4255
/rpt_family="(TGAA)n"
4277. 4368
/rpt_family="L2"
4693. 4822
/rpt_family="Achobo"
4842. 5142
/rpt_family="Alu"
5124. 5143
/rpt_family="(TAAA)n"
6061. 6192
/rpt_family="Achobo"
8244. 8813
/note="similar to EST BE985413 (NTID:gi0658696)"
8591. 9720
/note="CpG Island (%GC=75.5, o/e=0.90, #CpGs=161)"
8733. 8760
/rpt_family="GC_rich"
8766. 8800
/rpt_family="GC_rich"
8825. 8848
/rpt_family="GC_rich"
9243. 9289
/rpt_family="GC_rich"
10181. 10660
/rpt_family="MaLR"
10895. 10894
/rpt_family="Alu"
10995. 11281
/rpt_family="L1"
11268. 11296
/rpt_family="(TAAAAA)n"
11282. 11474
/rpt_family="Alu"
11449. 11544
/rpt_family="GA-rich"
11825. 12183
/rpt_family="CR1"
12210. 12312
/rpt_family="Achobo"
12313. 12651
/rpt_family="MaLR"
12652. 12797
/rpt_family="Achobo"
12854. 13010
/rpt_family="Achobo"
13012. 13390
/rpt_family="Achobo"
13355. 13387
/rpt_family="(TTG)n"
13391. 13681
/rpt_family="Alu"
13682. 13798
/rpt_family="Achobo"
13807. 14038

```

```

repeat_region
14022. 14071
/rpt_family="Alu"
14072. 14334
/rpt_family="GA-rich"
14308. 14478
/rpt_family="G-rich"
14490. 14534
/rpt_family="GA-rich"
14581. 14869
/rpt_family="Achobo"
15348. 15695
/rpt_family="Alu"
15745. 15832
/rpt_family="L2"
15925. 15970
/rpt_family="MaLR"
16115. 16455
/rpt_family="MaLR"
16567. 16660
/rpt_family="L2"
16661. 16852
/rpt_family="Alu"
16845. 16890
/rpt_family="(GGAGAA)n"
17631. 17661
/rpt_family="(CTG)n"
18253. 18341
/rpt_family="MIR"
18799. 19077
/rpt_family="Alu"
19077. 19189
/rpt_family="(TATATG)n"
19669. 19797
/rpt_family="ERVL"
19798. 20105
/rpt_family="Alu"

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent similarity: 100.000  Percent Identity: 100.000

```

```
alignment_block:
US-09-674-779-2 x AC068544/rev ..

```

```
Align seg 1/1 to reverse of: AC068544 from: 1 to: 159397

```

```

28 ProGlnValSerProileLysThrPro 36
|||||
15836 CCACAAAGTGAGCCCATCAAACTCCT 15810

```

```
seq_name: gb.htg:AC106164

```

```
seq_documentation_block:

```

```

LOCUS AC106164
DEFINITION Rattus norvegicus clone CH230-97117, *** SEQUENCING IN PROGRESS
ACCESSION AC106164
VERSION AC106164.1 GI:18138685
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```
REFERENCE
1 (bases 1 to 178928)

```

```

AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

```

Tue Sep 17 07:27:50 2002

```

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,C.D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabis,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzkert,M., Miner,G., Miner,Z., Mitchell,T., Newton,N.,
Morgan,M., Morris,S., Moser,M., Neal,D., Nickerson,E., Nwokkwo,S.,
Nguyen,A., Nguyen,N., Nguyen,N., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Ogih,M., Okuwona,G., Oragunye,N., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Peery,J., Perez,L., Perez,L., Rojas,A., Rojebokan,I., Roife,M.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Roife,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Woodson,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 178928)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUKA
Center clone name: CH230-97117
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findphraplist
Consensus quality: 145158 bases at least Q40
Consensus quality: 158004 bases at least Q30
Consensus quality: 167877 bases at least Q20
Estimated insert size: 133921; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 83 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 7669: contig of 7669 bp in length
* 7770 7769: gap of unknown length
* 7770 11596: contig of 3827 bp in length
* 11597 11696: gap of unknown length

```


variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known, annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

RP11-328K6 is from the library RP11-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-328K6 The true left end of clone RP11-203H2 is at 153730 in this sequence.

FEATURES

source

1. .198141
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-328K6"
/clone_lib="RP11-11.2"
53. .404

repeat_region

repeat_region
/note="MT1A1 repeat: matches 1. .365 of consensus"
547. .644
/note="TIGR2 repeat: matches 2398. .2500 of consensus"
1036. .1337
/note="AluSg repeat: matches 1. .303 of consensus"
1971. .2264
/note="AluX repeat: matches 1. .293 of consensus"
2482. .2708
/note="AluX repeat: matches 2. .225 of consensus"
2847. .3275
/note="TIGR2a repeat: matches 1. .433 of consensus"
3385. .3695
/note="AluSg repeat: matches 1. .309 of consensus"
4027. .4135
/note="L2 repeat: matches 2563. .2685 of consensus"
5269. .5327
/note="L2 repeat: matches 2364. .2419 of consensus"
6287. .6583
/note="AluSg repeat: matches 1. .297 of consensus"
6836. .6940
/note="L1ME3A repeat: matches 5996. .6103 of consensus"
7103. .7244
/note="MIR repeat: matches 20. .192 of consensus"
7359. .7629
/note="MIR repeat: matches 5. .268 of consensus"
7631. .7918
/note="AluSg repeat: matches 1. .288 of consensus"
7919. .7958
/note="20 copies 2 mer ca 100% conserved"
8561. .9002
/note="MIR repeat: matches 1. .448 of consensus"
9110. .9410
/note="AluSg repeat: matches 1. .301 of consensus"
9411. .9448
/note="19 copies 2 mer ta 100% conserved"
9604. .9689
/note="MIR repeat: matches 41. .126 of consensus"
9988. .10197
/note="L2 repeat: matches 2472. .2671 of consensus"
10438. .10865
/note="CpG island"
/evidence="not_experimental"

repeat_region

repeat_region
/note="L1MC5 repeat: matches 6437. .7589 of consensus"
12263. .13439
13548. .13696
/note="MER5B repeat: matches 2. .159 of consensus"
14380. .14549
/note="L2 repeat: matches 2534. .2690 of consensus"
14550. .14840
/note="AluSg1 repeat: matches 1. .288 of consensus"
14841. .15039
/note="L2 repeat: matches 2305. .2534 of consensus"
15098. .15448
/note="MT2CB repeat: matches 3. .411 of consensus"
15449. .15807
/note="THE1B repeat: matches 1. .364 of consensus"
15808. .15914
/note="MT2CB repeat: matches 362. .456 of consensus"
15927. .16206
/note="AluSg repeat: matches 1. .278 of consensus"
16208. .16494
/note="L2 repeat: matches 2129. .2419 of consensus"
16587. .16715
/note="L2 repeat: matches 2575. .2709 of consensus"
17070. .17215
/note="MIR repeat: matches 100. .252 of consensus"
17925. .17969
/note="L2 repeat: matches 2705. .2750 of consensus"
18038. .18356
/note="AluX repeat: matches 1. .312 of consensus"
18394. .18566
/note="FRAM repeat: matches -1. .171 of consensus"
18570. .18625
/note="28 copies 2 mer aa 75% conserved"
19223. .19310
/note="MT1J repeat: matches 103. .193 of consensus"
19877. .20039
/note="LTR33 repeat: matches 350. .519 of consensus"
20403. .20721
/note="AluX repeat: matches 1. .308 of consensus"
21963. .22029
/note="L1M4 repeat: matches 5721. .5794 of consensus"
22047. .22340
/note="AluY repeat: matches 1. .294 of consensus"
22864. .23282
/note="L2 repeat: matches 2027. .2491 of consensus"
23285. .23570
/note="MT1J repeat: matches 211. .514 of consensus"
23748. .24062
/note="MT1J repeat: matches 36. .389 of consensus"
24289. .24401
/note="MER30 repeat: matches 11. .130 of consensus"
24402. .24712
/note="AluY repeat: matches 1. .307 of consensus"
24713. .24814
/note="MER30 repeat: matches 130. .229 of consensus"
25096. .25391
/note="AluSg repeat: matches 1. .294 of consensus"
25554. .25849
/note="AluY repeat: matches 1. .295 of consensus"
26135. .26440
/note="AluSg repeat: matches 1. .312 of consensus"
27172. .27466
/note="AluSg repeat: matches 1. .296 of consensus"
27883. .27997
/note="L1M9 repeat: matches 6171. .6302 of consensus"
28577. .28610
/note="17 copies 2 mer tt 85% conserved"
28778. .28863
/note="MT1-INTERNAL repeat: matches 431. .525 of consensus"
28864. .29250
/note="MSTA repeat: matches 1. .426 of consensus"
29251. .29276
/note="MT1-INTERNAL repeat: matches 404. .431 of

```

Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA564014
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid: L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 209055 bases at least Q40
Consensus quality: 209344 bases at least Q30
Consensus quality: 209466 bases at least Q20
Insert size: 209790; sum-of-contents
Insert size: 194715; 2.8% error; agarose-fp
Quality coverage: 8.31x in Q20 bases; sum-of-contents Quality
coverage: 8.96x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..209790
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-564O14"
/clone_lib="RPC1-11.2"
1..209790
/note="assembly_fragment:04568"
BASE COUNT 58386 a 41949 c 43652 g 65793 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AL512373/rev ..
Align seg 1/1 to reverse of: AL512373 from: 1 to: 209790

16 MetLeuValAlaCysSerAlaProIle 24
|||||
57293 ATGCTGTTGCTTGCTGCGCCCATC 57267

seg_name: gb_pat:AX081147

seg_documentation_block:
LOCUS AX081147 40 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 3 from Patent WO0109335.
ACCESSION AX081147
VERSION AX081147.1 GI:13170041
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Thonnard,J.
TITLE Moraxella catarrhalis antigen basb120
JOURNAL Patent: WO 0109335-A 3 08-FEB-2001;
SmithKline Beecham Biologicals S.a. (BE)
FEATURES
Location/Qualifiers
1..40
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer"
BASE COUNT 17 a 4 c 7 g 12 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8

```

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AX081147 ..

Align seg 1/1 to: AX081147 from: 1 to: 40

1 MetLysAsnPheAsnGlnTyrPhe 8
|||||
16 ATGAAAAATTTAATCAACTTT 39

seq_name: gb_pat:AX081148

seq_documentation_block:

LOCUS AX081148 60 bp DNA linear PAT 27-FEB-2001

DEFINITION Sequence 4 from Patent WO0109335.

ACCESSION AX081148

VERSION AX081148.1 GI:13170042

KEYWORDS

SOURCE synthetic construct.

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 60)

AUTHORS Thonnard,J.

TITLE Moraxella catarrhalis antigen bas120

JOURNAL Patent: WO 0109335-A 4 08-FEB-2001;

SmithKline Beecham Biologicals s.a. (BE)

FEATURES

source 1..60
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="primer"

BASE COUNT 14 a 5 c 21 g 20 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AX081148/rev ..

Align seg 1/1 to reverse of: AX081148 from: 1 to: 60

243 HisValLeuProLysAsnLysLeu 250

|||||
60 CATGCTTTACCAAAAAATAAGCTA 37

seq_name: gb_sts:G22057

seq_documentation_block:

LOCUS G22057 430 bp DNA linear STS 31-MAY-1996

DEFINITION human STS WI-13488, sequence tagged site.

ACCESSION G22057

VERSION G22057.1 GI:1342383

KEYWORDS STS; STS sequence; primer; sequence tagged site.

SOURCE human STS derived from sequences in dbEST and the Unigene

collection.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 430)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Hudson,T.

JOURNAL Whitehead Institute/MIT Center for Genome Research; Physically

COMMENT Mapped STS
Unpublished

Contact: Thomas Hudson

Whitehead Institute/MIT Center for Genome Research

Whitehead Institute for Biomedical Research

9 Cambridge Center, Cambridge MA 02142 USA

Tel: 617 252 1900

Fax: 617 252 1902

Email: thudson@genome.wi.mit.edu

Primer A: TTTTCTTTTACACATACCAAAACA

Primer B: AAGTCTTTGGATGATTTTAAAGG

STS size: 135

PCR Profile:

Presoak:

Denaturation:

Annealing: 56 degrees C

Polymerization:

PCR Cycles: 35

Thermal Cycler:

Protocol:

Template: 10 ng

Primer: each 5 pM

dNTPs: each 4 nM

Taq Polymerase: 0.025 units/ul

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCl: 10 mM

pH: 9.3

Derived from dbEST (genbank accession R40672).

Location/Qualifiers

1..430

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="798.2 CR from top of Chr1 linkage group"

4..138

primer_bind 4..28

primer_bind complement(114..138)

BASE COUNT 115 a 81 c 84 g 146 t 4 others

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x G22057/rev ..

Align seg 1/1 to reverse of: G22057 from: 1 to: 430

24 IleProThrAsnProGlnValSer 31

|||||

374 ATACCCACNAATCCACAGTGAGT 351

seq_name: gb_sts:G61482

seq_documentation_block:

LOCUS G61482 650 bp DNA linear STS 30-MAR-2000

DEFINITION SHGC-85859 Human Homo sapiens STS genomic, sequence tagged site.

ACCESSION G61482

VERSION G61482.1 GI:6126651

KEYWORDS STS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 650)

AUTHORS Olivier,M. and Cox,D.R.

TITLE Unpublished, Olivier, M., Cox, D.R. (2000)

JOURNAL Unpublished

COMMENT

Contact: Michael Olivier, David R. Cox

Stanford Human Genome Center
Stanford University School of Medicine
4005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA
Tel: (650) 320-5800
Fax: (650) 320-5801

Email: olivier@shgc.stanford.edu
Primer A: TATCTCAGCAGTTTGGAGGGAG
Primer B: AGCTGAATCTGTCAACCTCGAG
STS size: 326
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplitaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES

Source 1..650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="13"
/clone_lib="Human"

STS 6..331

primer_bind 6..28

primer_bind complement(309..331)

BASE COUNT 185 a 109 c 144 g 212 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x G61482 ..
Align seg 1/1 to: G61482 from: 1 to: 650

8 PheIleThrLeuIleSerSer 15

|||||
100 TTTATTACCACTTTAATTCTTCA 123

seq_name: em_htgo_inv:AC085197

seq_documentation_block:

ID AC085197 standard; DNA; HTG; 854 BP.

XX AC AC085197;

XX AC AC085197.1

SV AC085197.1

XX 02-DEC-2000 (Rel. 66, Created)

DT 02-DEC-2000 (Rel. 66, Last updated, Version 1)

DE Giardia intestinalis clone KJ4544 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.

XX HTG: HTGS_PHASE0.

XX

OS Giardia intestinalis

OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.

XX [1]

RN 1-854

RX MEDLINE; 20389616.

RA McArthur A.G., Morrison H.G., Nixon J.E.J., Passananeck N.Q.E., Kim U.,

Hinkle G., Crocker M.K., Holder M.E., Farr R., Reich C.I., Olsen G.J.,

RA Alek S.B., Adam R.D., Gillin F.D., Sogin M.L.;

RT "The Giardia genome project database";

RL FEMS Microbiol. Lett. 189(2):271-273(2000).

XX [2]

RN 1-854

RA Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Fierro L.A.,

RA Alek S.B., Sogin M.L.;

RT Submitted (29-NOV-2000) to the EMBL/GenBank/DBJ databases.

RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,

RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA

XX

CC * NOTE: This record contains 1 individual

CC * sequencing reads that have not been assembled into

CC * contigs. Runs of N are used to separate the reads

CC * and the order in which they appear is completely

CC * arbitrary. Low-pass sequence sampling is useful for

CC * identifying clones that may be gene-rich and allows

CC * overlap relationships among clones to be deduced.

CC * however, it should not be assumed that this clone

CC * will be sequenced to completion. In the event that

CC * the record is updated, the accession number will

CC * be preserved. 854: contig of 854 bp in length.

CC 1

XX Key

EH Location/Qualifiers

EH 1..854

FT /db_xref="taxon:5741"

FT /organism="Giardia intestinalis"

FT /strain="WB-C6"

FT /clone="KJ4544"

XX

SQ Sequence 854 BP; 208 A; 200 C; 215 G; 228 T; 3 other;

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AC085197 ..

Align seg 1/1 to: AC085197 from: 1 to: 854

12 LeuIleSerSerMetLeuValAla 19

|||||

211 TTGATTTCTTCATGTTGTCGCA 234

seq_name: gb_ba:MSGLIPOPRE

seq_documentation_block:

LOCUS MSGLIPOPRE 900 bp DNA linear BCT 26-APR-1993

DEFINITION Mycobacterium intracellulare lipoprotein precursor (M143) gene,

complete cds.

ACCESSION L08229

VERSION L08229.1 GI:149971

KEYWORDS lipoprotein; precursor protein.

SOURCE Mycobacterium intracellulare (strain 13950) (library: lambda gtl1)

DNA.

ORGANISM Mycobacterium intracellulare

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE 1 (bases 1 to 900)
AUTHORS Nair, J., Rouse, D.A. and Morris, S.L.
TITLE Nucleotide sequence analysis and serologic characterization of a
27-kilodalton Mycobacterium intracellulare lipoprotein
JOURNAL Infect. Immun. 61, 1074-1081 (1993)
MEDLINE 93162797
FEATURES
    Location/Qualifiers
        1..900
        /organism="Mycobacterium intracellulare"
        /strain="13950"
        /db_xref="taxon:1767"
        /tissue_lib="lambda gt11"
        13..16
        25..813
        /standard_name="M143 lipoprotein precursor"
        /function="unknown"
        /note="lipoprotein consensus sequence is LSACGS in
        precursor peptide."
        /codon_start=1
        /transl_table=11
        /product="lipoprotein"
        /protein_id="AA25364.1"
        /db_xref="GI:149972"
        /translation="MSASCAVPLRFAVFAVAGATALSACGSSNKSSTSTSTST
        STSTVTSAPSPSTNAEAKVGLIASVAGNSIQVTQEDNATAVNETSATKITEAVPA
        GLPDVTGSGCLIVKPTGSGAPGQVTAAKVKISESVNGCTCPKPHSTPGGASSTPSPG
        SPSPAPAKPAWVDSVASVSGDPIINLTGDSGNITQITVTDVDDTKYTKQTANTEA
        IAPGKCLISARCTDSGALQATSIKURQAVDGKCGKPKPGGGG"
        repeat_region order(847..858,868..879)
        /rpt_type="inverted"
    BASE COUNT 172 a 339 c 284 g 105 t
    ORIGIN

alignment_scores:
    Quality: 8.00 Length: 8
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
    US-09-674-779-2 x MSGLIPOPRE rev ..
    Align seg 1/1 to reverse of: MSGLIPOPRE from: 1 to: 900
    16 MetLeuValAlaCysSerAlaPro 23
    |||||
    749 ATGCTCGTGCCTGCAGCGCCG 726
seq_name: em_htgo_inv:AC028827
seq_documentation_block:
ID AC028827 standard; DNA; HTG; 939 BP.
XX AC028827;
AC AC028827;
XX AC028827.1
XX 05-APR-2000 (Rel. 63, Created)
DT 05-APR-2000 (Rel. 63, Last updated, Version 1)
XX Giardia intestinalis clone HF2271 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
DE Giardia intestinalis clone HF2271 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX HTG; HTGS_PHASE0.
XX Giardia intestinalis
OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX [1]
RN 1-939
RA Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
Hinkle G., Holder M.E., Sogin M.L.;

```

```

RT "Giardia: a model for ancient eukaryotic genome analysis";
RL Unpublished.
XX [2]
RN 1-939
RP Hinkle G., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U.,
RA Crocker M.K., Holder M.E., Sogin M.L.;
RT ;
RL Submitted (01-APR-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.
CC * 1 939: contig of 939 bp in length.
XX Key Location/Qualifiers
FH source 1..939
FT /db_xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="HF2271"
XX
SQ Sequence 939 BP; 212 A; 240 C; 218 G; 268 T; 1 other;

alignment_scores:
    Quality: 8.00 Length: 8
    Ratio: 1.000 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
    US-09-674-779-2 x AC028827 ..
    Align seg 1/1 to: AC028827 from: 1 to: 939
    12 LeulleSerMetLeuValAla 19
    |||||
    889 TTGATTTCTTCATGTTGTCGCA 912
seq_name: em_htgo_inv:AC071017
seq_documentation_block:
ID AC071017 standard; DNA; HTG; 974 BP.
XX AC071017;
AC AC071017;
XX AC071017.1
XX 13-JUN-2000 (Rel. 64, Created)
DT 13-JUN-2000 (Rel. 64, Last updated, Version 1)
XX Giardia intestinalis clone EJ7174 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
DE Giardia intestinalis clone EJ7174 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX HTG; HTGS_PHASE0.
XX Giardia intestinalis
OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX [1]
RN 1-974
RA Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
Hinkle G., Holder M.E., Sogin M.L.;

```

Tue Sep 17 07:27:50 2002

```

RT "Giardia: a model for ancient eukaryotic genome analysis";
RL Unpublished.
XX [2]
RN 1-974
RP Eakin N.Q., Morrison H.G., McArthur A.G., Nixon J., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RL Submitted (06-JUN-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.
CC 1
XX 974: contig of 974 bp in length.
XX
FH Key Location/Qualifiers
FT source
FT 1. 974
FT /db_xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="EJ7174"
FT
XX
SQ Sequence 974 BP; 271 A; 231 C; 241 G; 230 T; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AC071017/rev ..
Align seg 1/1 to reverse of: AC071017 from: 1 to: 974

12 LeuIleSerMetLeuValAla 19
|||||
387 TTGATTTCTTCAATGTTGGTCGCA 364

seq_name: em_htgo_inv:AC060939
seq_documentation_block:
ID AC060939 standard; DNA; HTG; 984 BP.
XX
AC AC060939;
XX
SV AC060939.1
XX
DT 24-APR-2000 (Rel. 63, Created)
DT 24-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE Giardia intestinalis clone MJ4058 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW HTG; HTGS_PHASE0.
XX
OS Giardia intestinalis
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX
RN [1]
RP 1-984
RA Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;

```

```

RT "Giardia: a model for ancient eukaryotic genome analysis";
RL Unpublished.
XX [2]
RN 1-984
RP Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RL Submitted (20-APR-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.
CC 1
XX 984: contig of 984 bp in length.
XX
FH Key Location/Qualifiers
FT source
FT 1. 984
FT /db_xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="MJ4058"
FT
XX
SQ Sequence 984 BP; 247 A; 234 C; 256 G; 247 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AC060939 ..
Align seg 1/1 to: AC060939 from: 1 to: 984

12 LeuIleSerMetLeuValAla 19
|||||
264 TTGATTTCTTCAATGTTGGTCGCA 287

seq_name: em_htgo_inv:AC076436
seq_documentation_block:
ID AC076436 standard; DNA; HTG; 1002 BP.
XX
AC AC076436;
XX
SV AC076436.1
XX
DT 02-AUG-2000 (Rel. 64, Created)
DT 02-AUG-2000 (Rel. 64, Last updated, Version 1)
XX
DE Giardia intestinalis clone KJ4133 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW HTG; HTGS_PHASE0.
XX
OS Giardia intestinalis
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
XX
RN [1]
RP 1-1002
RA Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;

```


"Giardia: a model for ancient eukaryotic genome analysis";
Unpublished.

[2]
RN 1-1002
RA Kim U., Morrison H.G., McArthur A.G., Nixon J., Eskin N.O., Crocker M.C.,
RA Hinkle G., Holder M.E., Sogin M.L.;
RT ;

RL Submitted (01-AUG-2000) to the EMBL/GenBank/DBJ databases.
RL Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX

CC * NOTE: This record contains 1 individual
CC * sequencing reads that have not been assembled into
CC * contigs. Runs of N are used to separate the reads
CC * and the order in which they appear is completely
CC * arbitrary. Low-pass sequence sampling is useful for
CC * identifying clones that may be gene-rich and allows
CC * overlap relationships among clones to be deduced.
CC * However, it should not be assumed that this clone
CC * will be sequenced to completion. In the event that
CC * the record is updated, the accession number will
CC * be preserved.

CC 1 1002: contig of 1002 bp in length.

XX Key Location/Qualifiers
FH
FH
FT source 1. .1002
FT /db_xref="taxon:5741"
FT /organism="Giardia intestinalis"
FT /strain="WB-C6"
FT /clone="KJ4133"
XX

SQ Sequence 1002 BP; 245 A; 237 C; 248 G; 271 T; 1 other;

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AC076436 ..
Align seg 1/1 to: AC076436 from: 1 to: 1002

12 Leu1eserSerMetLeuValAla 19
|||||
403 TTGATTCTTCAATGTTGGTCGA 426

seq_name: gb_pl:ATH271473

seq_documentation_block:

LOCUS ATH271473 1226 bp mRNA linear PLN 25-JAN-2000
DEFINITION Arabidopsis thaliana mRNA for hypothetical protein related to
strictosidine synthase.

ACCESSION AJ271473

VERSION AJ271473.1 GI:6759490

KEYWORDS strictosidine synthase.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 1226)

Mahon,P.

REFERENCE Theses (2000) University of Cambridge

JOURNAL 2 (bases 1 to 1226)

AUTHORS Mahon,P.

REFERENCE Direct Submission

TITLE Submitted (18-JAN-2000) Mahon P., Biochemistry, Cambridge

JOURNAL University, Tennis Court Road, Cambridge, CB2 1QW, UNITED KINGDOM

COMMENT This sequence is from an unknown EST clone sent in error.

FEATURES

source

1. .1226 Location/Qualifiers
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
<1. .1061
/function="related to strictosidine synthases"
/note="ORF"

CDS

/codon_start=3
/product="hypothetical protein"
/protein_id="CAB69786.1"
/db_xref="GI:6759491"
/translation="HASAIIGDNLKGTGKLVQDAKTIPLVDGPESLEFDPQEGPY
VGVTDRILKWRGEELGWDFAYTSPHRDNCSSHEVVPSCGRLGLSFERKTGDLYIC
DGYFGVMKVGPEGLAELVDEAGRKVMEANOGDIDDEEDIFVFNDSPTYHERDVF
YVLSGKTGVGRVIRYDMKKKRAKIMDKLRPLNGLSKNGSVFVTCESSTNTCHRIW
VKPKSGTNEVFATLPGSPDNIRPTPTGDFWALHCKKNLFRAYLIHTVWGRFFMT
MKMERVIHFMMNGKPRGHIVVVKLSGTGEILLEDESGKTVKYVSEAYETKDGKLVIG
SVYWPVWVLDTISVYDSI"
polyA_site 366 a 225 c 312 g 323 t
BASE COUNT
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x ATH271473/rev ..

Align seg 1/1 to reverse of: ATH271473 from: 1 to: 1226

144 ProAlaAsnThrGlnIleArgSer 151
|||||
1020 CCGGCCAATACACAGATCCGATCC 997

seq_name: gb_pat:E03308

seq_documentation_block:

LOCUS E03308 1262 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA encoding recombinant streptokinase.

ACCESSION E03308

VERSION E03308.1 GI:2171525

KEYWORDS JP 1992011892-A/1.

SOURCE synthetic construct.

ORGANISM synthetic construct.

REFERENCE 1 (bases 1 to 1262)

AUTHORS Fujii,S., Katano,T., Majima,E., Ogino,K., Ono,K., Sakata,Y. and
Uenoyama,T.

TITLE PROTEIN OF STREPTOKINASES, CORRESPONDING GENE, CORRESPONDING
PLASMID RECOMBINANT, CORRESPONDING CHARACTER TRANSFORMANT AND
PRODUCTION

JOURNAL Patent: JP 1992011892-A 1 16-JAN-1992;

OSUKA PHARMACEUTICAL FACTORY INC

OC Artificial gene

PN JP 1992011892-A/1

PD 16-JAN-1992

PF 06-JUL-1990 JP 1990179851

PR 11-JUL-1989 JP 89P 179432, 27-NOV-1989 JP 89P 307957, PR

11-APR-1990 JP 90P 96830

PI FUJII SETSUO, KATANO TAMITAKA, MAJIMA EIJI, OGINO KOICHI, PI

ONO KENJI,

PI SAKATA YASUYO, UENOYAMA TSUTOMU

PC C12N15/58,C12N1/21,C12N9/770,(C12N1/21,C12R1:19),(C12N9/70, PC

C12R1:19);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC *source: clone=PSKX;

PH Key Location/Qualifiers

FT 5'UTR 1..11

FT CDS 12..1256

FT mat_peptide 12..1253

FT 3'UTR 1257..1263

FEATURES

source

1..1262

/organism="synthetic construct"

/db_xref="taxon:32630"

BASE COUNT 365 a 350 c 262 g 285 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x E03308/rev ..

Align seg 1/1 to reverse of: E03308 from: 1 to: 1262

192 GluLeuGlnAsnArgLeuCysGln 199

|||||

393 GAGTTACAGAACCGCTCTTTGTGAC 370

seq_name: gb_pl:AF370181

seq_documentation_block:

LOCUS AF370181 1407 bp mRNA linear PLN 30-APR-2001

DEFINITION Arabidopsis thaliana unknown protein (F24I3.90/AT3g57010) mRNA,

complete cds.

AF370181

AF370181.1 GI:13877836

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 1407)

Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,

Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,

Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,

Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,

Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,

Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,

Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and

Theologis,A.

Full length cDNA of gene F24I3.90/AT3g57010 (GI:6911871)

Unpublished

2 (bases 1 to 1407)

Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,

Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,

Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,

Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,

Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,

Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,

Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and

Theologis,A.

Direct Submission

Submitted (18-APR-2001) Plant Gene Expression Center, 800 Buchanan

Street, Albany, CA 94710, USA

RIKEN Genomic Sciences Center (GSC) members carried out the

collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN

Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,

Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,

Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the

sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,

Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,

Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L.,

Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E.,

Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J.,

Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to

this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)

contributed equally to this work as PIs.

FEATURES

source

1..1407

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAFL05-09-N10 (R10213)"

/note="This clone is in a modified pBluescript vector

(Lambda ZAP) as a XhoI/SstI insert.

ecotype: Columbia"

1..1407

/gene="F24I3.90/AT3g57010"

1..97

/gene="F24I3.90/AT3g57010"

98..1228

/gene="F24I3.90/AT3g57010"

/codon_start=1

/evidence="experimental

/product="unknown protein"

/protein_id="AAK43996.1"

/db_xref="GI:13877837"

/translation="MPISRRVLPTTAAPVILAVLCFFWSSIIIGPNLKGKHLQD

AKTIPVDGPESLEFPDQGPVYDGRILKWRGELGWVDFAVTSRDNCSH

EWPCSGRPLGSLFERKTGLYICDGVYKMGVPEGLGELVVDVDEGRKVMFANQG

IDDEEDIFVNDSSDYHFHDFYVLSGTGKVRVIRYDMKKKAKVIMDKLRLPNG

LALSKNGSVFYTCSTNICHRIWVGPKSGTNEVFATPGSPDNIRRTPTGDFWAL

HCKNLETRAVLIHTWGRFMTMTMETVIHFVNGKGRHGIIVKLSGTGETLEILE

DSEKTVKYVSEAYETDKLWIGSVYVPAVWLDTSVYDSI"

misc_difference 522

/gene="F24I3.90/AT3g57010"

/note="compared to genomic

acid sequence difference"

/replace="c"

1228..1407

/gene="F24I3.90/AT3g57010"

BASE COUNT 411 a 260 c 343 g 393 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AF370181/rev ..

Align seg 1/1 to reverse of: AF370181 from: 1 to: 1407

144 ProAlaAsnThrGlnIleArgSer 151

|||||

1187 CCGGCCAATACAGATCCGATCC 1164

seq_name: gb_ov:IPJ00267

seq_documentation_block:

LOCUS IPJ00267

DEFINITION Ictalurus punctatus mRNA for Oct1 transcription factor.

ACCESSION AJ000267

VERSION AJ000267.1 GI:2739211

KEYWORDS

SOURCE

ORGANISM

Ictalurus punctatus

Tue Sep 17 07:27:50 2002

```

LOCUS               AK022355 Homo sapiens cDNA FLJ12293 fis, clone MAMMA1001815.
DEFINITION          AK022355 Homo sapiens cDNA FLJ12293 fis, clone MAMMA1001815.
ACCESSION            AK022355
VERSION              AK022355.1 GI:10433734
KEYWORDS              oligo capping; fis (full insert sequence).
SOURCE                Homo sapiens Mammary gland cDNA to mRNA, clone_lib:MAMMA1
                     Clone:MAMMA1001815.
ORGANISM              Homo sapiens
                     Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE            1 (sites)
AUTHORS              Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
                     Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
                     Tanase,T., Nomura,Y., Togiya,S., Komai,F., Hara,K., Takeuchi,K.,
                     Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
                     Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuho,Y. and Oshima,A.
                     NEDO human cDNA sequencing project
                     NEDO human cDNA sequencing project supported by Ministry of
                     International Trade and Industry of Japan; cDNA full insert
                     sequencing; Research Association for Biotechnology; cDNA library
                     construction, 5'- & 3'-end one pass sequencing and clone selection;
                     Helix Research Institute (supported by Japan Key Technology Center
                     etc.) and Department of Virology, Institute of Medical Science,
                     University of Tokyo.
FEATURES             Location/Qualifiers
                     1..2077
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MAMMA1001815"
                     /clone_lib="MAMMA1"
                     /tissue_type="Mammary gland"
                     /note="cloning vector: pHE18SFL3"
BASE COUNT          630 a 452 c 395 g 600 t
ORIGIN
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
  US-09-674-779-2 x AK022355/rev ..
  Align seg 1/1 to reverse of: AK022355 from: 1 to: 2077
      211 GlyLeuGlyLeuTyAlaThrGly 218
      |||||||
      937 GGACTAGGACTTTATGCCACTGGA 914
seq_name: gb_ba:AB010463
seq_documentation_block:
  LOCUS              AB010463
  DEFINITION          Vibrio parahaemolyticus gene for NorM, complete cds.
  ACCESSION            AB010463
  VERSION              AB010463.1 GI:3298335
  KEYWORDS              NorM.
  SOURCE              Vibrio parahaemolyticus (strain:AQ3334) DNA.
  ORGANISM              Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
  REFERENCE            1 (bases 1 to 2243)
  AUTHORS              Tsuchiya,T.
  TITLE              Direct Submission
  JOURNAL              Submitted (17-JAN-1998) Tomofusa Tsuchiya, Okayama University,
LOCUS               US-09-674-779-2 x FCCPKS19 rev ..
DEFINITION          US-09-674-779-2 x FCCPKS19 from: 1 to: 2010
ACCESSION            US-09-674-779-2 x FCCPKS19
VERSION              US-09-674-779-2 x FCCPKS19
KEYWORDS              192 GluLeuGlnAsnArgLeuCysGln 199
                     |||||||
                     1093 GAATTACAAAATCGGTTATGTCAG 1070
seq_name: gb_pat:E12464
seq_documentation_block:
  LOCUS              E12464
  DEFINITION          Capsid protein gene of feline calicivirus.
  ACCESSION            E12464
  VERSION              E12464.1 GI:3251297
  KEYWORDS              JP 1997000267-A/2.
  SOURCE              unidentified.
  ORGANISM              unclassified.
  REFERENCE            1 (bases 1 to 2034)
  AUTHORS              Kenjo,T., Maeda,T., Yokoyama,N. and Toya,Y.
  TITLE              RECOMBINED FELINE HERPESVIRUS 1 TYPE CAPABLE OF MUTATING THYMIDINE
                     KINASE GENE AND VACCINE CONTAINING THE SAME
  JOURNAL              Patent: JP 1997000267-A 2 07-JAN-1997;
                     KYORITSU SHOJI KK
  COMMENT              PN JP 1997000267-A/2
                     PD 07-JAN-1997
                     PF 23-JUN-1995 JP 1995179609
                     PI KENJO TAKESHI, MAEDA TAKESHI, YOKOYAMA NAOAKI, TOYA YUKINOBU
                     PC C12N15/09,A61K39/245,C12N5/10,C12N7/00,C12N15/09,C12R1/92);
                     CC strandedness: Double;
                     CC topology: Linear;
                     FH Key
                     FH Location/Qualifiers
                     1..2034
                     /organism="Feline calicivirus" FT CDS
                     /product="capsid protein".
FEATURES             Location/Qualifiers
                     1..2034
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT          524 a 493 c 448 g 569 t
ORIGIN
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
  US-09-674-779-2 x E12464/rev ..
  Align seg 1/1 to reverse of: E12464 from: 1 to: 2034
      192 GluLeuGlnAsnArgLeuCysGln 199
      |||||||
      1108 GAATTACAAAATCGGTTATGTCAG 1085
seq_name: gb_pr:AK022355
seq_documentation_block:

```

Faculty of Pharmaceutical Sciences; 1-1-1 Tsushima-naka, Okayama,
 Okayama 700-8530, Japan
 (E-mail:tsuchiya@phsant.pharm.okayama-u.ac.jp,
 Tel:81-86-251-7957, Fax:81-86-251-7957)
 2 (sites)
 Morita,Y., Kodama,K., Shiota,S., Mine,T., Kataoka,A., Mizushima,T.
 and Tsuchiya,T.
 Norm, a putative multidrug efflux protein, of *Vibrio*
parahaemolyticus and its homolog in *Escherichia coli*
 Antimicrob. Agents Chemother. 42 (7), 1778-1782 (1998)
 98325450

FEATURES
 Source Location/Qualifiers
 1. .2243
 /organism="Vibrio parahaemolyticus"
 /strain="AQ3334"
 /db_xref="taxon:670"
 833. .2203
 /gene="norm"
 833. .2203
 /gene="norm"
 /codon_start=1
 /transl_table=11
 /product="Norm"
 /protein_id="BAA31456.1"
 /db_xref="GI:3298336"
 /translation="MRYKEASSLIKLPVLIASVAQTGMGFVDTVMAGVYTOTDM
 AAVSASSIWLPSILFGTGLMALVPVVAQLNGSARREKIPFIEQGVVLALLISIP
 IGVLLQTOFIIQMDVEAMADKTGYTHAVFAVPAFLFOTLRSFTGMSLTKPKAM
 VTGFGILLNIPNLFVYVFGAPGLGCGVATTVYVWFALLALIVMTSSRLK
 SINVEGVEHKPKQAVRLKLGPEVAALFEVFLFAVALLVSPGLPIIVAAHQA
 INFSLVFLMPSVCAVSIKRGHRLGEENVDAVSRVGMVGLALATITAITVL
 SRELLAELTYNNPEVITLALQVLLFAVYQCTDAVQVIAAGALRGYKDMRAIFNRTFI
 AWIILGFTGYILGRTDWIVPEMGAQSGFWLGIITGLTAAALMLGVRLRWHRQEPDQV
 LNFSLQ"

BASE COUNT 527 a 537 c 503 g 676 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AB010463 ..

Align seg 1/1 to: AB010463 from: 1 to: 2243

137 LeuYsSerArgGlyLeuPro 144
 |||||
 598 CTGAAGTCCGCGCATTTTGCT 621

seq_name: gb_v1:CVXCAPS

seq_documentation_block:
 LOCUS CVXCAPS 2493 bp ss-RNA linear VRL 02-AUG-1993
 DEFINITION Feline calicivirus capsid protein gene (put.), complete cds.
 ACCESSION M32819
 VERSION M32819.1 GI:323874
 KEYWORDS capsid protein.
 SOURCE Feline calicivirus
 ORGANISM Viruses; ssRNA positive-strand viruses, no DNA stage;
 Caliciviridae; Vesivirus.
 1 (bases 1 to 2493)
 Neill,J.D., Reardon,I.M. and Heinrikson,R.L.
 Nucleotide sequence and expression of the capsid protein gene of
 feline calicivirus
 J. Virol. 65, 5440-5447 (1991)
 91374597

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Draft entry and computer-readable sequence for [Unpublished (1990)]
 kindly submitted
 by J.D.Neill, 14-MAR-1990. Pioneer Hi-Bred International, 7300 NW

62nd Ave., Johnston, IA 50131.
 Location/Qualifiers
 1. .2493
 /organism="Feline calicivirus"
 /db_xref="taxon:11978"
 127. .2133
 /note="capsid protein ORF (put.); putative"
 /codon_start=1
 /protein_id="AAA42925.1"
 /db_xref="GI:323875"
 /translation="MCSTCANVLYKYYDMDPHIKLVINPNKFLHVGFCNDPLMCCVPEL
 LPEFGTMDQDSPLOVYILEILGDDSSSTHEAIDPVVPMHNDGACKIFQPHGVL
 MHLICKVAEGWDPNLPRLRLEADGSIITPQGTWGVGVAEPNAGMSTDAADATGK
 SVDSWEAFSEFTSNVWSTSETQCKILFKOSIGPLLPYLTHLAKLVAGSWGVDVR
 FTSISGSGVGGKAAIVVPPGIDPVQSTSMQYVPHVLEDAKQVEFVIFSIDPLKSTLY
 HLMSDTDTSLVIMVNDLINPYANDSNSGCVITVETKPGDFKHLKPKPGSMLTH
 GSIPDLIPKSSSLMIGNFSDITDFVIRPFVQANRHFDFNQETAGWSTPRFRPIT
 ITTSVKESAKLGIVATDYIVPGIDGWPDTTIPGELVPVGDYAITNGTNNIDITAAQ
 YDAATEIRNTNFRGMYICGSLQRAWGDKISNTAFITVTVDGAKLIPSNITDQTKI
 AVFQDTHANKHVQTSIDVNSOILHTSROLNHYLLSPDSFAYVRIIDNSGWSWFDIGID
 HKNSIKLGYVIRSIDVNSOILHTSROLNHYLLSPDSFAYVRIIDNSGWSWFDIGID
 NDGSEFVGVSISGKLEFPLTASYMGIQIAKIRLASINSVMTKL"
 2130. .2450

FEATURES
 Source Location/Qualifiers
 1. .2493
 /organism="Feline calicivirus"
 /db_xref="taxon:11978"
 127. .2133
 /note="capsid protein ORF (put.); putative"
 /codon_start=1
 /protein_id="AAA42925.1"
 /db_xref="GI:323875"
 /translation="MCSTCANVLYKYYDMDPHIKLVINPNKFLHVGFCNDPLMCCVPEL
 LPEFGTMDQDSPLOVYILEILGDDSSSTHEAIDPVVPMHNDGACKIFQPHGVL
 MHLICKVAEGWDPNLPRLRLEADGSIITPQGTWGVGVAEPNAGMSTDAADATGK
 SVDSWEAFSEFTSNVWSTSETQCKILFKOSIGPLLPYLTHLAKLVAGSWGVDVR
 FTSISGSGVGGKAAIVVPPGIDPVQSTSMQYVPHVLEDAKQVEFVIFSIDPLKSTLY
 HLMSDTDTSLVIMVNDLINPYANDSNSGCVITVETKPGDFKHLKPKPGSMLTH
 GSIPDLIPKSSSLMIGNFSDITDFVIRPFVQANRHFDFNQETAGWSTPRFRPIT
 ITTSVKESAKLGIVATDYIVPGIDGWPDTTIPGELVPVGDYAITNGTNNIDITAAQ
 YDAATEIRNTNFRGMYICGSLQRAWGDKISNTAFITVTVDGAKLIPSNITDQTKI
 AVFQDTHANKHVQTSIDVNSOILHTSROLNHYLLSPDSFAYVRIIDNSGWSWFDIGID
 HKNSIKLGYVIRSIDVNSOILHTSROLNHYLLSPDSFAYVRIIDNSGWSWFDIGID
 NDGSEFVGVSISGKLEFPLTASYMGIQIAKIRLASINSVMTKL"
 2130. .2450

FEATURES
 Source Location/Qualifiers
 1. .2493
 /organism="Feline calicivirus"
 /db_xref="taxon:11978"
 127. .2133
 /note="capsid protein ORF (put.); putative"
 /codon_start=1
 /protein_id="AAA42926.1"
 /db_xref="GI:323876"
 /translation="MNSTLGLIDVTWTKGAOOIELDKAALGOORELALORNLRQ
 ALNNOVEQFNLLKQVQGPQSVLARAAGFRVDPYSYTNQNFYDDQLNAIRLSYRN
 LFKN"

BASE COUNT 666 a 593 c 531 g 703 t
 ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x CVXCAPS/rev ..

Align seg 1/1 to reverse of: CVXCAPS from: 1 to: 2493

192 GluLeuChnAsnArgLeuCysGln 199
 |||||
 1216 GAATTACAAATCGGTATGTCAG 1193

seq_name: gb_ov:AF072657

seq_documentation_block:
 LOCUS AF072657 2927 bp mRNA linear VRT 17-DEC-1998
 DEFINITION Danio rerio HLH-containing transcription factor (Coe2) mRNA,
 complete cds.
 ACCESSION AF072657
 VERSION AF072657.1 GI:3273890
 KEYWORDS zebrafish.
 SOURCE Danio rerio
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 2927)
 Bally-Cuif,L., Dubois,L. and Vincent,A.
 Molecular cloning of zcoe2, the zebrafish homolog of Xenopus Xcoe2
 and mouse EBF-2, and its expression during primary neurogenesis
 Mech. Dev. 77 (1), 85-90 (1998)
 99002949

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 2927)
 Bally-Cuif,L., Dubois,L. and Vincent,A.
 Direct Submission
 Submitted (16-JUN-1998) CNRS URA 1414, Ecole Normale Supérieure, 46
 rue d'Ulm, Paris 75005, France

/note="acute myeloid leukemia patient"

1. .3500
/gene="AF9q34"
148. .3345
/gene="AF9q34"
/note="MLL fusion partner"
/codon_start=1
/product="nGAP-like protein"
/protein_id="AAK50336.1"
/db_xref="GI:14009346"

/translation="MPRLKESRSHESLLSPSSAVEALDLSMEEVWIKPVHSSILGQD
YCEVTTSSGSCFCRSARABDKMENLRVHNDKNSRRVHILKLVIAKDL
AKKYLCELDLDVLTARTGLKLDNFGEHEEHLPLRTVTVHLYRTDKKK
KERNYGLVSLVPAASVAGQVFKWYVPTFNAGGKPGPMIRIKARYOTLILPM
EMVKEFAETHYILGLCAALEPILSARTKEEMASALVHLQSGIKVDELTDLMSSE
VDCGNDELIFRENTLATKALEEYKLVGKQLODAGEFIKALIESDCEVDPK
CSAADLPEHQGNLKWCELAFCIIINSYCVFPRELKEVFASWQECSSRGPDISERL
ISASLEFLCPLAIMSPSLFNLOEYPPDRTARTLTIAKVTQNLANFAKPGSKEEYM
SFNQLEHEWNTMQRFLEISNPETLSNTAGFEYIDLGRELSSLHLLWEAVSOLE
QSVKLGPLPLRLDRVHTALSTPGSGQLPGTNDLASTPGSGSSISAGLQNGKSLM
LSGLIDTFLPPTPENKDLFFVTRSSGVQVPSPARSSYSSEANPDLOMANGRAGOT
VDQARTLDGAGSPAGDVLPTDQAAAOLVAGVAPARATPVNLAGLATVRAGOT
PTTPGTSEAGPRQLPLSPQNPVYQMAAGLPLSPRLGDSSEGHSSLSHSNSE
ELAAKLGFSFAEELARRQGLARQMLTERGGQVTPRQNSAGPQRIDQPPP
PPPPPPAPRGTPPNLSTLOYPPSSCTLASPDWVGVSTRLRQSSSKGDSPE
LKPRVHKGQSPVSPNADLRTAAWLLTNAOLLEDEGLGDPDRHRLERLRQDQKD
EKDLAVLQDKLRISTKKELEYETLFKQETTKLVLEYQARLEGEERLRQDQKD
IQMKIISRLMSVEELKDDHAAVDSKRIIDAEKRIASLDAAANRLMSALTO
LKERYSMOARNGISPTNPTKLQITENGEPFRNSNC"

BASE COUNT 769 a 1109 c 1057 g 565 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AY032952/rev ..
Align seg 1/1 to reverse of: AY032952 from: 1 to: 3500

101 LeuLeuThrAlaArgSerIrrp 108
|||||
2821 CTCTTAACAACGCGGCTTCATGG 2798

seq_name: gb_vi:FCL3E

seq_documentation_block:
LOCUS FCL3E 3516 bp RNA linear VRL 07-FEB-1999
DEFINITION Feline calicivirus (FCV) genome, 3'-end, RNA polymerase and capsid precursor protein.

ACCESSION D90357.1 GI:221264
VERSION D90357.1
KEYWORDS RNA polymerase; capsid precursor protein.
SOURCE Feline calicivirus F4, CDNA to genomic RNA, clone pFCV119.
ORGANISM Feline calicivirus

REFERENCE 1 (bases 1 to 3516)
AUTHORS Tohya,Y., Taniguchi,X., Takahashi,E., Utagawa,E., Takeda,N., Miyamura,K., Yamazaki,S. and Mikami,T.
TITLE Sequence analysis of the 3'-end of feline calicivirus genome
JOURNAL Virology 183 (2), 810-814 (1991)
MEDLINE 91306470
COMMENT Submitted (18-Mar-1991) to DDBJ by: Mikami

Takeshi

Department of Veterinary Microbiology
Faculty of Agriculture, The University of Tokyo
1-1-1 Yayoi, Bunkyo-ku
Tokyo 113
Japan

gene
CDS

Location/Qualifiers
1. .2927
/organism="Danio rerio"
/db_xref="taxon:7955"
1. .2927
/gene="Coe2"
417. .2156
/note="similar to Xenopus laevis Xcoe2 and Mus musculus EBF2"
/codon_start=1
/product="HbH-containing transcription factor"
/protein_id="AAC96103.1"
/db_xref="GI:3273891"

/translation="MFESQDHSIRITITELKVRTFEEMDPVRSWVRNVGVIDANIAAQ
SGVALSRAHEKOPPNLRKSNFFHVLALYDRNGQPVVEVTSFVDFEODKTEKT
NGTHVQLQLLYSNGVTRTQDLYARLIDSVTKQPISEYQGNKNPEMCRVLLTHEVMCS
RCKKSCGNRETSPDVTIDREFLKEFLKCNCLKTAGNPRDMRRFQVVLSTVVC
VDGVLALISDNMETHVHNSKHGRSRMRDNETVENNMYATPCIKAIQSPGWTGGA
MVIYIGENFFDGLQVFGVMSLWSELITPHAIRVQTPPHIIPGVVETLVSYSKQFCK
GAPGRFYIALNEPTIDYGVPRPHSQQAAMPSPVHSGVMGLSSYPTQLGVSGIGEPGQ
SNOMLLKRAADIAEALYSVPRPSSTPQQAAYSGMGMSYGAVPMSSLVGSGPGFNASAP
TSOGYTRNSSSLSIPGYPSSSTPQQAAYSGMGMSYGAVPMSSLVGSGPGFNASAP
NSSPYAIMPSPGSSSSSLLPFSSPSSSTKQSAFAPVLRPQGFPHPSAKTSGGT
SFRATMLGLVPPM"

BASE COUNT 811 a 646 c 658 g 812 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AF072657 ..
Align seg 1/1 to: AF072657 from: 1 to: 2927

98 MetSerGlnLeuThrAla 105
|||||
1453 ATGAGCCCACTATTGACACGGCT 1476

seq_name: gb_pr:AY032952

seq_documentation_block:
LOCUS AY032952 3500 bp mRNA linear PRI 08-MAY-2001
DEFINITION Homo sapiens nGAP-like protein (AF9q34) mRNA, complete cds.

ACCESSION AY032952
VERSION AY032952.1 GI:14009345
KEYWORDS human.
SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 3500)
AUTHORS von Bergh,A.R.M., Wijers-Koster,P.M., Groot,A.J., Kluin,P.M. and Schuurin,E.
TITLE A novel Ras GTPase Activating Protein (RasGAP) gene is fused to MLL in acute myeloid leukemia with t(9;11)(q34;q23)

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3500)
AUTHORS von Bergh,A.R.M., Wijers-Koster,P.M., Groot,A.J., Kluin,P.M. and Schuurin,E.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2001) Department of Pathology, Leiden University Medical Center, Albinusdreef 2, Leiden 2333 ZA, The Netherlands

FEATURES
source

1. .3500
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q34"

Phone: 03-3812-2111 x5396
 Fax: 03-5689-7346.
 Location/Qualifiers
 1. .3516
 /organism="Feline calicivirus"
 /db_xref="taxon:11978"
 <1. .1146
 /note="RNA polymerase"
 /codon_start=1
 /protein_id="BAAL4370.1"
 /db_xref="GI:221265"
 /translation="LDLSSKKLATOGIALPHEHYTICLDELPRIEKMOEGKRMW
 GCDVAVTAAAFKGVSDAITANHQYGIQVGNMDSVPSVEVLYORIKSAAKVEAVD
 FSNWDSQSPRVSAASIDILFYDSPIVDSAAITLKSPPIAENGVAVKYAGSLPS
 GMYLSTVNSLNHCWYGCATLQSLERQIPVTWNLFSSDMYIGDDGYMFPPTMFA
 SVSDOIFGNLSAYGLKPRVDSKGAIESIDPSEVFLKRTITPTNGIRGLDRSSI
 IRQFYIIGENSDDKTPTKIDTIDPSRQQLWNAQVLSQHGVEFYFNKVLKLAAMRAVE
 YEGHLKPPSSSALLEHYNSQFNGEARSQINMSDVTALHCDVFEV"
 1149. .3155
 /note="capsid precursor protein"
 /codon_start=1
 /protein_id="BAAL4371.1"
 /db_xref="GI:221266"
 /translation="MCSSTCANVLYQWDPHFRLIINPKFLPIGFCDFNLMCCYPDL
 LPEFTGVDCOSPLQIYLESILGDEWSTHEAIDPSVPMHWDAGKTFQPHRGVL
 MHHLIGEYAKAMDNLPLFLREADDGSIITPEOGTAVGGVIAEPSAQMSAADMSGK
 SVDSWEAFFSEHTSVANMSTSETGKILFKOSLGPLLNPYLEHLSKLYVAMSGSIEVR
 FSISSGVFGKLAALVVPFGVQVSTSMQLQPHVLFDAQVEPVFTIPDLRSLY
 HWSDDTTSVIMVINDLINPYANDNSNGCIQVTFETKPGDFKFLHLLKPPGSLVTH
 GSIFDLIPKSSIGWIGNRYWTDITDFVIRPFVQANRHFDFNOETAGWSTPRFRPIT
 ITISEKNSKIGIGVADYIIPGIPDGWPTTADKLIIPAGDYSITTGRENIDKTAQA
 YDTAAVKNNTNFRGMVIGSLQRAWGDKKISNTARTTIAIRDGNEIKPSNTIDMTKL
 AVYDQTHVEQVQSDDTLALLGVYTGIEEAGSNDRVVRSVLEPAGARGGNHPID
 YKNSIKLGVYRSIDVENSQILHTSRQSLNHLPLPPDSFVAVYRIIDNSGWFIDIG
 SEGSPFVGSDIGKLEFPLSASYNGIQLAKIRLASNRSMTKL"
 3152. .3472
 /note="short open reading frame"
 /codon_start=1
 /protein_id="BAAL4372.1"
 /db_xref="GI:221267"
 /translation="MNSILGLIDVTNTVGVKGOOIELDKAALGOORELALORMSLDRQ
 ALNQVQSFKNLLEQRVHPIQSVRLARAAGFRVDPYSYTNQNFYDQLNAILRSYRN
 LFRN"
 192 GluLeuGlnAsnArgLeuCysGln 199
 2238 GAATTACAAATCGGTATGTCAG 2215
 seq_name: gb_pr:AB051530
 seq_documentation_block:
 LOCUS AB051530 3790 bp mRNA linear PRI 07-FEB-2001
 DEFINITION Homo sapiens mRNA for KIAA1743 protein, partial cds.
 ACCESSION AB051530
 VERSION AB051530.1 GI:12698030
 KEYWORDS
 SOURCE Homo sapiens cDNA to mRNA, clone:pj01380.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (sites)
 Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.
 Prediction of the coding sequences of unidentified human genes.
 XIX. The complete sequences of 100 new cDNA clones from brain which
 code for large proteins in vitro
 DNA Res. 7 (6), 347-355 (2000)
 21082932
 2 (bases 1 to 3790)
 REFERENCE
 Ohara,O., Nagase,T. and Kikuno,R.
 Direct Submission
 Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,
 Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba -
 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
 URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
 Fax:81-438-52-3914)
 FEATURES
 Location/Qualifiers
 1. .3790
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="pj01380"
 /note="vector:pBluescriptII SK plus"
 1. .1807
 /gene="KIAA1743"
 <1. .1807
 /gene="KIAA1743"
 /note="Start codon is not identified."
 /codon_start=2
 /product="KIAA1743 protein"
 /protein_id="BAB21834.1"
 /db_xref="GI:12698031"
 /translation="NLANFAKFGSKSEYKFMNQFLEHWTNNQRFLEISNPETLSN
 TAGPEYIDIGRELISLHLLWEAVSQLEQISIVSKGLPLPRILRDVHTALSTPGSGQL
 PGTNDLASTPGSGSSSTISAGLQKRWIENDLSGLIDFTRLPSPPTPENKDLFFVTRSGV
 QPSPARSSSYSEANPDLOMANGKSLSDVLDQADARTLDGEAGSPAGPDVLPDQGA
 AAQLVAGWPARATPVNLGLATVVRAGQTPPTTSEGAGRPQLLAPLSTFQNPYQW
 AAGLPSPRGIDSGSEHSSLSSESEAAKLSFSTAEEELARRPGLARRQ
 MSLTEKGGQTPVQNSAGPQRIDQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
 TLASADPDWGLGPTSLRLQSSSSKSLSPKRAVHKQGPSVPSNADLRTAAWLLTM
 NAOLEDGLGPDPPHRRDLRSKDELQSAEKDLAVLDQKLRLSTKKLEVEYTLFKQE
 ETTQKLVLEQVQARLEGEERLRQEQEDQIQKGIISRLMSVEEELKKDHAEMQAAYD
 SKOKIIDAQEKRIASLDANARLMSALTQIKESMH"
 BASE COUNT 776 a 1181 c 1121 g 712 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-674-779-2 x AB051530/rev ..
 Align seg 1/1 to reverse of: AB051530 from: 1 to: 3790
 101 LeuLeuThrAlaArgSerTrp 108
 1370 CTTCTAACAACTGCGGTCATGG 1347
 seq_name: gb_ro:AF236130
 seq_documentation_block:
 LOCUS AF236130 6128 bp mRNA linear ROD 22-AUG-2001
 DEFINITION Rattus norvegicus DOC2/DAB2 interactive protein mRNA, complete cds.
 ACCESSION AF236130
 VERSION AF236130.1 GI:15277524
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 6128)

BASE COUNT 932 a 828 c 788 g 968 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-674-779-2 x FCL3E/rev ..
 Align seg 1/1 to reverse of: FCL3E from: 1 to: 3516
 192 GluLeuGlnAsnArgLeuCysGln 199
 2238 GAATTACAAATCGGTATGTCAG 2215
 seq_name: gb_pr:AB051530
 seq_documentation_block:
 LOCUS AB051530 3790 bp mRNA linear PRI 07-FEB-2001
 DEFINITION Homo sapiens mRNA for KIAA1743 protein, partial cds.
 ACCESSION AB051530
 VERSION AB051530.1 GI:12698030
 KEYWORDS
 SOURCE Homo sapiens cDNA to mRNA, clone:pj01380.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Primates; Catarrhini; Homnidae; Homo.

FEATURES
 source
 CDS
 /organism="Feline calicivirus"
 /db_xref="taxon:11978"
 <1. .1146
 /note="RNA polymerase"
 /codon_start=1
 /protein_id="BAAL4370.1"
 /db_xref="GI:221265"
 /translation="LDLSSKKLATOGIALPHEHYTICLDELPRIEKMOEGKRMW
 GCDVAVTAAAFKGVSDAITANHQYGIQVGNMDSVPSVEVLYORIKSAAKVEAVD
 FSNWDSQSPRVSAASIDILFYDSPIVDSAAITLKSPPIAENGVAVKYAGSLPS
 GMYLSTVNSLNHCWYGCATLQSLERQIPVTWNLFSSDMYIGDDGYMFPPTMFA
 SVSDOIFGNLSAYGLKPRVDSKGAIESIDPSEVFLKRTITPTNGIRGLDRSSI
 IRQFYIIGENSDDKTPTKIDTIDPSRQQLWNAQVLSQHGVEFYFNKVLKLAAMRAVE
 YEGHLKPPSSSALLEHYNSQFNGEARSQINMSDVTALHCDVFEV"
 1149. .3155
 /note="capsid precursor protein"
 /codon_start=1
 /protein_id="BAAL4371.1"
 /db_xref="GI:221266"
 /translation="MCSSTCANVLYQWDPHFRLIINPKFLPIGFCDFNLMCCYPDL
 LPEFTGVDCOSPLQIYLESILGDEWSTHEAIDPSVPMHWDAGKTFQPHRGVL
 MHHLIGEYAKAMDNLPLFLREADDGSIITPEOGTAVGGVIAEPSAQMSAADMSGK
 SVDSWEAFFSEHTSVANMSTSETGKILFKOSLGPLLNPYLEHLSKLYVAMSGSIEVR
 FSISSGVFGKLAALVVPFGVQVSTSMQLQPHVLFDAQVEPVFTIPDLRSLY
 HWSDDTTSVIMVINDLINPYANDNSNGCIQVTFETKPGDFKFLHLLKPPGSLVTH
 GSIFDLIPKSSIGWIGNRYWTDITDFVIRPFVQANRHFDFNOETAGWSTPRFRPIT
 ITISEKNSKIGIGVADYIIPGIPDGWPTTADKLIIPAGDYSITTGRENIDKTAQA
 YDTAAVKNNTNFRGMVIGSLQRAWGDKKISNTARTTIAIRDGNEIKPSNTIDMTKL
 AVYDQTHVEQVQSDDTLALLGVYTGIEEAGSNDRVVRSVLEPAGARGGNHPID
 YKNSIKLGVYRSIDVENSQILHTSRQSLNHLPLPPDSFVAVYRIIDNSGWFIDIG
 SEGSPFVGSDIGKLEFPLSASYNGIQLAKIRLASNRSMTKL"
 3152. .3472
 /note="short open reading frame"
 /codon_start=1
 /protein_id="BAAL4372.1"
 /db_xref="GI:221267"
 /translation="MNSILGLIDVTNTVGVKGOOIELDKAALGOORELALORMSLDRQ
 ALNQVQSFKNLLEQRVHPIQSVRLARAAGFRVDPYSYTNQNFYDQLNAILRSYRN
 LFRN"

AUTHORS Zhi.W., Tseng,C.-P. and Hsieh,J.-T.
TITLE A novel Ras GTPase activating protein that interacts with
DOC-2/DAB2: a downstream effector leading to the suppression of
prostate cancer
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6128)
AUTHORS Zhi.W.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-2000) Urology, UT Southwestern Medical Center,
5323 Harry Hines Blvd, Dallas, TX 75390-9110, USA
FEATURES Location/Qualifiers
source 1. .6128
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
CDS 458. .3448
/note="DIP1/2"
/codon_start=1
/product="DOC2/DAB2 interactive protein"
/protein_id="AAK93947.1"
/db_xref="GI:15277525"
/translation="MENLRVHPKNSRRVEHILKLVIEAKDLPAKKYLCGLCL
DVLVYATGKLTNDVFWGEHFEHFNPLPRTIVIVHLYRETDKKKERSYGLIYS
LPAASVAGQFVEKVPVTPNPKGGKGPMPKIRKQITITILPMEMIKFAEHIT
NHVLTGCAALEPTLSAKTEEMASALVHIOSTGVKDFDLTDMSEVDRGCDNEHLI
FRNTLATGIEEYILKGVHKLQDALCEFIKALYESDENCEVDPKSGAADLPFHOG
NLKMCCELAFCIKINSYVTPRELKEVFASWQECSSRGPRDISERLISASLFLFLC
PALMSSELNLQIYDDPTARTLTIAKVTGNLANFAKFGSKSEYFMFGNGLGLPL
TNMORLELISNETLUSNAGFEGYIDLGRELSHLLWEAVSQDQSIKVLGLPL
RIURDHTALSPGQGLPNDLASTPGSSSVSTGLQKSVIENDLSGLIDFTRLP
SPTPNKDLFFVTRSSGVQSPARSSYSEANEPDLOMANGSKVLSMDQDARTLDG
EAGSPVPALPADGVQVATQVAVGPARAAPSGLATVTRAVPTTPTGSEAGP
GRQLLAFLSPFNPTVMAAGLPLSPRLGDSGSEHSSLSHNSSEELAAAKLGSE
STAAELARPEGLABROMSLTEKGGQVTPRONAGPORRIDOPPPPPPPAPRG
RTPPTMLSTQVPRSSGLASPDWAGPTELQOQSSSKGDSPELKPRLRHQGP
SPVSNALDQTAWLITMNAQLLEDEGLGPDPPHRLRKSKEELSOAEKDLAVLDKL
RISTKLEIYETFKCOETTKLVLEYQARLEEGEERLRQOQDQVDMKGIISRLM
SVBEELKDHAEQMAVDQKQIIDAQEKRIASLDAAANRLMSALTQLKERYSMRARN
GVSTPNTKLIQTENGFEFRNSSC"
BASE COUNT 1389 a 1791 c 1713 g 1235 t
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x AF236130/rev ..
Align seg 1/1 to reverse of: AF236130 from: 1 to: 6128
101 LeuLeuThrThrAlaArgSerTrp 108
2924 CTCTTAACAACCTGCGGCTTCATGG 2901
seq_name: gb_pat:AX251205
seq_documentation_block:
LOCUS AX251205 6291 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 173 from Patent WO0168912.
ACCESSION AX251205
VERSION AX251205.1 GI:15984628
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 6291)
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 173 20-SEP-2001;
Epigenomics AG (DE)

FEATURES Location/Qualifiers
source 1. .6291
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 1501 a 210 c 1626 g 2954 t
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x AX251205/rev ..
Align seg 1/1 to reverse of: AX251205 from: 1 to: 6291
34 LysThrProSerValLeuIleThr 41
1884 AAAACACCATCCGTTTAATTACT 1861
seq_name: gb_vi:FCU13992
seq_documentation_block:
LOCUS FCU13992 7677 bp RNA linear VRL 17-APR-1998
DEFINITION Feline calicivirus CFI/68 RNA helicase/cysteine
protease/RNA-dependent RNA polymerase polyprotein precursor and
capsid protein precursor, genes, complete cds; and unknown gene.
ACCESSION U13992
VERSION U13992.1 GI:3056875
KEYWORDS Feline calicivirus.
SOURCE Feline calicivirus.
ORGANISM Viruses: ssRNA positive-strand viruses, no DNA stage;
Caliciviridae; Vesivirus.
REFERENCE 1 (bases 1 to 7677)
AUTHORS Neill,J.D.
TITLE Complete nucleotide sequence of feline calicivirus strain CFI/68
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 7677)
AUTHORS Neill,J.D.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1994) VCR, National Animal Disease Center, P. O.
Box 70, Ames, IA 50010, USA
REFERENCE 3 (bases 1 to 7677)
AUTHORS Neill,J.D.
TITLE Direct Submission
JOURNAL Submitted (17-APR-1998) VCR, National Animal Disease Center, P. O.
Box 70, Ames, IA 50010, USA
REMARK Sequence update by submitter
COMMENT On Apr 17, 1998 this sequence version replaced gi:537254.
FEATURES Location/Qualifiers
source 1. .7677
/organism="Feline calicivirus"
/strain="CFI/68"
/specific_host="Felis domesticus"
/db_xref="taxon:11978"
5'UTR 1. .19
CDS 20. .5308
/codon_start=1
/product="RNA helicase/cysteine protease/RNA-dependent RNA
polymerase polyprotein precursor"
/protein_id="AAC13992.1"
/db_xref="GI:3056876"
/translation="MSQTLSEFLVTKHNVRKDFVRSVKLTLLARRDLQYFNKLSRSMR
AEACPSASYDVCNCTSDIPDDSGSTELIPSWEEVTKSTYLSLSEDSDELCPD
DIANVAAHIRKAIISTQSHSPANSCKEQLTSLVMAEAMLPQRSASIPLHQHOAAR
LEWRKFFSKPLDFELRGVSKDILQITAIWKILERKACYCKSYGQWFTAKQKLR
EWRSESNTLKLPLGAFIDGLRFMTVDNPYMGFLPKLIGLIKPLNLAITDNHETL
SGWITLTALMELYNITECTIDMTSLTAFYDKATKFKYSHVKALFTGTFEDVS
NSFWMAAAAILCYLTGTLIPNNGRFLKICLLVRAITLVLSGLIATQKLAAMPATWNSE

STVNELSARTVAISELNPNPTTSDTESVERLELEAKILHEEIKVHTLNPIMQSNLIL
RNLMTLDGVITCCNKRKAJAKKQVPCVYILTPGPGCGKTTAAQALAKLSDEPSV
INLVDHDTVTGNEVCILDEDSKDYANFVGMVNSAPMVLNCDMLNKGKLF
SKYILMTSNETPVPKSKRAGATRRVYTHDVATLVESHKRRPAGTAPRSCYKKNF
SHLSIAKAGCWCKEYVLDKPGHQHSTKAPPTFLNIDSLAQTMKQDFALKNAFE
AVGSCSEHYGFVCOQSEVETVRLLNARMLNATFTVCVGLSEASNSVGTATHLVP
DEPNKGRFVSRNEASALEGNCVOTALGCMNSKDLTHLCHFTKGTIVNDSVBL
DELPAHQHVTVNSVFLAWARHSLTGTQFOAIRAAYDVLHVDPKVPAMLRHMD
TSFSEHVVTOFIPGGVILVESCGRARIALGNNVIRAGVTAIPGCGVRLMGLSA
QTPWSEILSELGLKINSVKVYLLILALSIAFRFPKTEAKGKTKSIGPYR
GRGVALTDEDEKHEHNAKRLDLSVEFLMLRHAALGADTDVAFKFRSNNRSR
LADDEDTVYIGKGVHEKIRTNLRAVDGTYDVFAESGPGCAKHNKNAISGTV
CGEHGKGVHMGHYASVAVHVGKDSFFLGERIFDLKNGEFCFCCPSTLPSAAPP
FSGRPTDGPWSPATKPKPYSTSGKIVGCFATTSTHETPGDGLPYIDONGRVT
GLHTGSGPKTPSAKLVVYVHIDMKTKVTAQYDVKPDLSVKGVLQCLDEIRIL
PKGTRLVSPAHLDEFECSHOPASLGSDPCPKSLTAIVDSLSKPYCVVNGPHD
ILHRVOKMLDHLGSGFVPMNLSSTMSLAFHKLNDHSTSCGYLGGRRKDHMGVEDP
KALLDLSKKWLTATQGIAPHEYTIGLKDRLPIEKVQEGKRMWCGDVGATVCA
AAGVSDAITANHOYGPVQGINMDSPEALFQRIKSAKVFADVSKWDTQSPR
VSAASIDILRYFSRTPIDVATNLKSPPIAVNGVAVKVSGLPGLMPLTSVNSL
NHCLYVGCALQLSLEARNVPTWNLSTFDMVTVGDDVYVMTYASISDQIFANLS
AYGLKPTRVDKSVGSIPIIDPNSVFLKRTITRTPOGIRGLDRESSILROYIKEN
TDNKEPKTIDPMRGOQLWACLYASQHGIDFYNKYVLAEKAVETGHLHLEPPSY
STALEHNSQFNGVEARTDQIDTSGMAALHCDVFEV
5311..7317

CDS

/codon_start=1
/product="capsid protein precursor"
/protein_id="AACI3993.1"
/db_xref="GI:537256"

/translation="MCSTCANVLYKVDWDHPIKLVNPKFLHVGFCNDPLMCCYPEL
LPEGTWDCDSPLQVLESILGDDWSTHEAIDVPPVPMHDEAGKIFQPHPGVL
MHLICVAGSDPNLFLREADDGSIITPEQGTVMGVGVAEPNAQMSTAADMATGK
SVSEWFAFSEHTSNVNSTSETQKILFKQSLGILNPLYLTHLAKLYVAMSGSDVR
FSTSGSVFGGKLAIVVPPGIDPQVSTMLQVPHVLFDAQVPEVFESTPDLRSTLY
HMSDITTLVIMVNDLNPVANDSNSSCIVTVETKPGDFKHLKPPGSMILTH
GSPDLIPKSSILIGNRFMSDITDPIRPFVQANRHFDFQETAGNTPRFRPIT
ITSVKESAKLGVGADYIVGIPGWPDTITIPGELVPGVDIATNGNNDITTAQ
YDARTNNTNFRGMTICGSLQWAGDKKISATFTTGTVDGAKLIPSTIDQTKI
AVQDTHANKHVSDDTLALLGTSIGGEAIGADRVRVRSVLPERGARGNHPFI
HKNSIKIGYVIRSDVNSQILHTSRLSLNHYLLSPDSFVAVYRIDSNGSWFDIGID
NDGFSFVGSSIGKLEPPLTASYMGILAKIRLASNTRSVMTKL"
7314..7634

CDS

/codon_start=1
/product="unknown"
/protein_id="AACI3994.1"
/db_xref="GI:537257"

/translation="MNSILGLDVTNTTICKAQOIELDKAALGOORELALQRMNLDRO
ALNNQVEQFNKLEQRVQGPISQVRLARAAGPRVDPYSYTNQNFYDQLNATRLSYRN
LFKN"

BASE COUNT 2141 a 1767 c 1677 g 2092 t

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x FCU13992/rev ..

Align seg 1/1 to reverse of: FCU13992 from: 1 to: 7677

192 GluLeuGlnAsnArgLeuCysGln 199
|||||
6400 GAATACAAAATCGGTATATGAC 6377

seq_name: gb_vi:FCLF4

seq_documentation_block:

LOCUS FCLF4

DEFINITION Feline calicivirus genomic RNA for non-structural proteins,
complete cds.

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE

REFERENCE
AUTHORS

TITLE
JOURNAL

FEATURES
source

CDS

D31836
D31836.1 GI:499895
non-structural protein.
Feline calicivirus (strain:F4) cDNA to genomic RNA, clone:clones
pFCV1119, 148 and 2081.
Feline calicivirus
Viruses; ssRNA positive-strand viruses, no DNA stage;
Caliciviridae; Vesivirus.
1 (sites)
Oshikamo, R., Tohya, Y., Kawaguchi, Y., Tomonaga, K., Maeda, K.,
Takeda, N., Utagawa, E., Kai, C. and Mikami, T.
The molecular cloning and sequence of an open reading frame
encoding for non-structural proteins of feline calicivirus F4
strain isolated in Japan
J. Vet. Med. Sci. 56 (6), 1093-1099 (1994)
95210407
2 (bases 1 to 7681)
Tohya, Y.
Direct Submission
Submitted (15-JUN-1994) Yukinobu Tohya, Kagoshima University,
Faculty of Agriculture, Department of Veterinary Microbiology,
1-21-24 Koorimoto, Kagoshima, Kagoshima 890, Japan
(Tel:0992-85-8725, Fax:0992-85-8725)
Location/Qualifiers
1..7681
/organism="Feline calicivirus"
/strain="F4"
/db_xref="taxon:11978"
/clone="clones pFCV1119, 148 and 2081"
20..5311
/note="ORF1 encoding the non-structural proteins"
/codon_start=1
/product="non-structural proteins"
/protein_id="BAA06622.1"
/db_xref="GI:1060879"

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0

BASE COUNT 2046 a 1810 c 1777 g 2048 t

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x FCLF4/rev ..

Align seg 1/1 to reverse of: FCLF4 from: 1 to: 7681

192 GluLeuGlnAsnArgLeuCysGln 199

|||||

6403 GAATTACAAAATCGGTATGTCAG 6380

seq_name: gb_ba:AF400582

seq_documentation_block:

LOCUS AF400582

DEFINITION Acinetobacter sp. ADP1 mismatch repair protein (mutS), 7-Fe

ferredoxin (fdx), and O-methyltransferase-like protein genes,

complete cds; and unknown genes.

ACCESSION AF400582

VERSION AF400582.1 GI:15217080

KEYWORDS J. Bacteriol. 183 (23), 6822-6831 (2001)

SOURCE Acinetobacter sp. ADP1.

ORGANISM Acinetobacter sp. ADP1

Bacteria: Proteobacteria; gamma subdivision; Moraxellaceae;

Acinetobacter.

1 (bases 1 to 8367)

Young, D.M. and Ornston, L.N.

Functions of the Mismatch Repair Gene mutS from Acinetobacter sp.

Strain ADP1

J. Bacteriol. 183 (23), 6822-6831 (2001)

2 (bases 1 to 8367)

Young, D.M. and Ornston, N.

Direct Submission

Submitted (18-JUL-2001) Molecular, Cellular, and Developmental

Biology, Yale University, P.O. Box 208103, New Haven, CT

06520-8103, USA

FEATURES

Location/Qualifiers

1..8367

/organism="Acinetobacter sp. ADP1"

/strain="ADP1; B0413"

/db_xref="taxon:62977"

complement(12..725)

/note="ORF5"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK92492.1"

/db_xref="GI:15217081"

/translation="MTLTQPLSNVNNFIDLNRKAEFAFEVFRKTTITANGTVN

EVERGPDNLITFNYPGWKSKVILVETLDGEVLLGRKTAGGRYNKLFREHVDVT

TISHVHANLGNATQHTLPIRYVPQRFHLKELPVIIDRROAEVDFILDQLKVDI

ETRALEANGATVWGRGLOETAEFILLLEGAQIQLLAELIGSREYGGVHLQWQ

KMSGLTERAKSLGLPLANN"

complement(739..1797)

/note="ORF4"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK92493.1"

/db_xref="GI:15217082"

/translation="MSLVNNKNSKEFVKNILSEIWTTRCPVPTATGISUKRGLTVK

KEKRGIDVAVLQDARKGLSVHFDHLPALFREGGNVPALAAKSEGAPSLIGLTIWI

DMQVLIIVRPSDIISQPEHLLNAKVALPEADKRAKSTFRAMSLQIGKALQLANLSE

RVNTEVPERAGRNKENDASQYWSGIDALITGOVDVYVKGASAVEAATKRGKLVGI

LDLQDPDITSRKNTPRTIIVHEDLLEDHFVDVDFLEELFDADWAQDHLDEVLTI

LQETIAGASVATAYRNNFHSLSLSTERLELLKQKDFLWLHGLFLERDVINDNW

VDRPLLEVLKRRQNKL"

2592..5237

/gene="mutS"

2592..5237

/gene="mutS"

gene

CDS

CDS

/note="Muts"
/codon_start=1
/transl_table=11
/product="mismatch repair protein"
/protein_id="AAK92494.1"
/db_xref="GI:15217083"
/translation="MNSTETMADLSSYTPMMQYFKVLEHQHALLFYRMGDFVELEFF
DDARKAAKFLGTLTHRGKANGEPIMAGVPYHAAEGYLAFLRVAGQTVLAEQVGG
ENAGSRCKAPMERKVRIITPGTITDDALLGYSQSNLVALCICQNKIGIALLDLSAS
IFKQQHEFKEQELYIELARLPSEIIVVDELDVQNTLEQIKKQIEGSIITKRPNVDFN
LNNAKTLCDFQGVTTLSGFGIDHPLAKAAAAALHYAKETOKTALPHRSIOLEOS
TDFALDPIITRNLEIIDPLFEHGTSLPNLINDSCVLTGSRVAGTAPRDLVQLRQACQIFPL
ARLDATEQLLIGYHESVPLVLEIGTIDVLSRVAGTAPRDLVQLRQACQIFPL
RHALLPAIQTKSSKLLNDELHDLGDFKDLHQLLSAIVNPPVLLRDNVIAEGYDAE
LDELKIRDHAGQFLVLEIEKERSSGIPTLKIGYNRVSGYIIELTRAQEAQAPAYXI
RROTLKNAERYITPELKSFEKDKVLSSESARALAKELLFESLLETLKKNIANLQWSSA
IAQIDLANFAHQARLQWNRPDSPSEIGIKIHGRHPVEALSKTPTVNDTFLDSQ
RMALITGNMGKSTFMKQFALISLACGYSYPAKSAKLGPIIDRTFRIGSADSL
TKSTFMVEMTETSQILHATHQSLVLMDEVGRGTSTYDGLSLAWACVLDITKRIKCL
CLFATHYFELTGLSESADINYHVTAKEMNGNLILLHKVQHPASQSHGLQVAKLGI
PASVTKKAKRLKILEKQOQYFOTAVQNDLFAIPELPQIDETKIEIKTSPVLDHLK
OLDVDSLTPROALEALYQKQOLEH"

5493..5942

/note="ORF1; similar to PAD-binding domain of AppA redox

regulator from Rhodobacter sphaeroides"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK92495.1"

/db_xref="GI:15217084"

/translation="MAYVRLLYVSTAKQHEIKHDLMEILTTAIRPNSINKIKCVLY
YNGYFVQCLEGEKVIYDFVYRIIKDSRHNOCILSKAVNELLFORNSMKFAPLN
KKLRDFFQOYHIDEFNPYLLTHQTTEPFIELLSQEPHQEVDPYPHSS"

6267..6590

/gene="fdx"

6367..6590

/note="fdxA; similar to AvfD1 from Azotobacter vinelandii"

/codon_start=1

/transl_table=11

/product="7-Fe ferredoxin"

/protein_id="AAK92496.1"

/db_xref="GI:15217085"

/translation="MTFVVTENCICKKYODCVVDCFYEGPNFLVINPDECIDCA
LCEPCPANAIFSEDELPEGQVFIELNADLSQKWENTIQIGDQADREWNKPKDKL
QYLEK"

6717..7439

/note="ORF2"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAK92497.1"

/db_xref="GI:15217086"

/translation="MSQFLKGVLAISIIPIMVGGCTAPQKKKILDLPLKGRHRIHPQ
ETVIERPVKVTNPNSYLNWLNQNSRIQIEYQFSLRNDVANIVPNYELLKRTARD
WOKGRSEVAVPNRELNNAIPTLRVLKVAAILTDFEVSVDLPINQAGGAN
SRHLFNSAIDFRIGPEIIPQODYAFIENTKFKCQFWNQHGQSLNMGGLYSSQIH
IDTQGYRTWGPDRHSRNSMCN"

7593..8261

/note="ORF3; member of O-methyltransferase family"

/codon_start=1

/transl_table=11

/product="O-methyltransferase-like protein"

/protein_id="AAK92498.1"

/db_xref="GI:15217087"

/translation="MOQLMTDIDHYTESNLIPEDSNLKFALQNTDOQFPDHLAVAPN
QGMLLQMLTOMNOCKRVLEGTAYSTIWLAKALPKDGLTMTIEGRDTHAALAQONI
DHANLPVNIELKGRAADILSKIDTDTDFEFDIFIDADKQSYPEYDLSLHSGGT
VIVLDNVIRAGALIDLDNORPSIEGIRLTFEALQNHPLLISCTALQTVGSKGDDGFAL
AIVK"

2656 a 1606 c 1445 g 2660 t

BASE COUNT

ORIGIN

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-674-779-2 x AF400582 ..
  Align seg 1/1 to: AF400582 from: 1 to: 8367

158 LeuAsnGlnCysAlaGlyGlyAla 165
|||||
7170 TTAATCAGTGTGCTGCTGCGCG 7193

seq_name: gb_ba:AE002287
seq_documentation_block:
LOCUS AE002287 10780 bp DNA linear BCT 26-MAY-2000
DEFINITION Chlamydia muridarum, section 19 of 85 of the complete genome.
ACCESSION AE002287 AE002160
VERSION AE002287.2 GI:8163162
KEYWORDS
SOURCE
ORGANISM
Chlamydia muridarum.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
1 (bases 1 to 10780)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
20150255
10684935
2 (bases 1 to 10780)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7190237.
Location/Qualifiers
1..10780
/organism="Chlamydia muridarum"
/db_xref="taxon:83560"
/notes="synonym: Chlamydia trachomatis MoPn"
complement(65..364)
/gene="TC0200"
complement(65..364)
/gene="TC0200"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF73538.1"
/db_xref="GI:8163163"
/translation="MFKSKPKKKKCCMMWLVFGGILATLLTSLFVPSKGIQIRKK
LRKVTSGTKGKRALRNKSHHREFAEQTKTLAKNISKEIQDTQSIIDSRRD"
complement(384..746)
/gene="TC0201"
complement(384..746)
/gene="TC0201"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11

```

```

/product="conserved hypothetical protein"
/protein_id="AAF39073.1"
/db_xref="GI:7190238"
/translation="MTPYLVEITLVCCILLSGLMWLIFAFAGILLKLKHKIMKHVSRL
ADLFNFEARLAPFLGKKFVNLWLRKKYNNRKISQIAEELNDEDESDSCSNRLIFG
VKLAALALCAWLFIKAKD"
991..2370
/gene="TC0202"
991..2370
/gene="TC0202"
/notes="similar to GB:L12968 SP:P31120 GB:U01376 PID:290443
PID:606114; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoglucomutase/phosphomannomutase family
protein"
/protein_id="AAF39074.1"
/db_xref="GI:7190239"
/translation="MTCDSVPLFGTDGIRGRANFEPMTVETSVLGLKAVAGVLLERHS
GHRVVGKDRSLSGTGNFALIGTSMGIELMLGPIPTPGVAFITRAYRADAGIM
ISASHNPYRDNGIKIFSSDFKIGQAVEERIEAMTASKSFGLGPDHAGVGNKRKVKDA
MGRIYAKATPKGRTLKLGRIVLDCAGAAAYRVAHSVFEIDAEVICYGCEPGCN
INEGCGALWPSIIQKAVIEHADVGLDGDGRLIMVDEKGIHVDGMLLSICASDL
KRRQALPENRVATVNTFNGVLRILIESLGIOVTISPVGDRHVLOHMLTKATLGEOG
GHMIFLDVNTGDIIVSALQVLRIMIESESTLSDLTACIVKSPQSLINVPVAKVPLD
SLSNVQVLOEYKEILGDSGRILLRYSGTENICRVWVEGTRKHQVDSLAKTIVDVVEA
EIGGAEISE"
2382..4202
/gene="TC0203"
2382..4202
/gene="TC0203"
/notes="similar to PID:1184044 SP:Q56213; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucosamine--fructose-6-phosphate
aminotransferase (isomerizing)"
/protein_id="AAF39075.1"
/db_xref="GI:7190240"
/translation="MCGIFGYLQQRNAVPLVLEGLSKLEYRGYSAGATAALTAKHFLV
EKSQVPSQLCSKSVSSMHSQVAIGTETWATGFEFRNAPHVDMVESCALVHNGII
ENFOTREDLSKGVESDSTDEVTIVQLFASRYRETGDVLQVQSWTLKQGSFACA
LVHQDHPILLICATYEVSLILGEEVFISDVHAFKYSQITLASGLAVLRIG
RPVEIYFELSRIOKEVRCIDHAGSLDKGFDYVLMKEIYEQPEVERILHFVCEEN
GFAESFLKFESEGTIESHIIVACGSSYHAGCLAKYVIESMVSIPYVETASEFRQRP
YIAKSLAILISQGETADTIALNFEKRLDEVRLGICNVGRSVLASRVDHCVFIDIA
GLEVGASTAKTAQLLVLLILGLKLSQROEISKQDLMOAQVGLRELPRLTRFLDIA
SIHWRCSQSKETSPFLGRFRFYPICMEAAKLKETAIVYEAANAYPAGEMKHGFIALI
QEGTPVIYCGDPFVYTKTIGAIMEVKARKAYVIALARESNOIDIAAASVDEQIYIPDSH
DLAAPILFAIAGQIMAYTMALQKGTEDVRPRNLAKSVTVE"
4306..5502
/gene="TC0204"
4306..5502
/gene="TC0204"
/notes="similar to GB:M23240 SP:P18199 PID:148089 GB:U00096
PID:1736569; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Mtr/TnaB/Tyro permease family protein"
/protein_id="AAF39076.1"
/db_xref="GI:7190241"
/translation="MINKMLGGALIVAGTTIGAGLVAVPIATSEGGFLPTLLVWSW
LIIVASGVCFLEVLTLHARKNNVNSMAECTLGYSKSVIMLVLYLLFYSLIVAVFC
DGGNILLRMVMSCRWDPTWIRHAMPVIFEFALFPLMAKTSIVDOCHNRVFLGLGISF
AMFCYFGFLPKTKTELLYRSSWGGTLAGFPFLFAGFQNVVPTLIHYTMKRNVRDVKKA
IVIGSFPLVLIWIWEAIVLGAVPVSLEQAKMEGWTAIGALQALKCSAFYTAGEFF
ECTFALISFIVGLGKDFIDAFQWDEKKRKLFIPLVFFVPLVWAVFPGVIVLKL
ECTGALGETIVLGVFVPLVMWKGKRYGKRYGQRIPLPGKGLIIVMSGLVLMVLIIV
QKFLGY"
5699..6886
/gene="TC0205"
5699..6886
/gene="TC0205"

```


Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

FEATURES

```

source
    Location/Qualifiers
    1..10986
    /organism="Xylella fastidiosa 9a5c"
    /db_xref="taxon:160492"
    /clone="9a5c"
    complement(91..1602)
    /gene="XF1282"
    complement(91..1602)
    /note="similar to GI16458686 (percent identity: 33 %/query
    alignment coverage: 91.1 %/subject alignment coverage:
    91.6 %); identified by sequence similarity; putative; ORF
    located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="carboxypeptidase related protein"
    /protein_id="AAF84091.1"
    /db_xref="GI:9106271"
    /translation="MAMIPRIALLSLVSSVPLAYATATPPKPKDRTTTPATLTPAL
    PANTQIRQIRNRLISYKAIIGALPVDDKCKTIGEVMTAYIMKGNRPVTFALN
    GPCASSVYLMGAIGPKVVFCTQGTSPDPAIHDNPGTWLDFDLPIDFPGF
    SRARIEGEAKOFFPKADYDLSRTIDWLVAAGRLLSKRLVGVSGYGGFRAPRIA
    HTLQTMQIALNLGLVLYSLNETLDNDSVPLAWIMTLPSTAAHLERQHQLSEAK
    MREVIETRGDYASTLMKGRSDPAATMLQOVVHLTGLDPTVRRAGVLTQAVLR
    EVFRDTELGSRDYDINVTALDPPFDSEQRTPDLDLSIIAPTPTAMVDFRTAGWK
    VHERYQALNDVNRQWNEELAKGSVTOLRQSVADPKLRLVLIAGHWNDLSCPFGS
    VLTIDQIPKMGSDPTRIQVREYPGGHMFYNRPDQSQSFQDQVQMYEHQDTP"
    1613..1828
    /gene="XF1283"
    1613..1828
    /note="XF1283"
    /note="hypothetical protein; identified by sequence
    similarity; putative; ORF located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAF84092.1"
    /db_xref="GI:9106272"
    /translation="MYRNSSRGEQWYQKLLKLEAMSSCGWRAFGVGGLCVVRLYRDD
    REVVVDDIVAVCYRSASVTRMFRSID"
    1821..1979
    /gene="XF1284"
    1821..1979
    /note="XF1284"
    /note="hypothetical protein; identified by sequence
    similarity; putative; ORF located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAF84093.1"
    /db_xref="GI:9106273"
    /translation="MIDPNALWILHGDGALPVSVDVWVFRGCCIVQRASLLLSPEAVV
    VLLALSME"
    complement(2061..2231)
    /gene="XF1285"
    complement(2061..2231)
    /note="XF1285"
    /note="hypothetical protein; identified by sequence
    similarity; putative; ORF located using Glimmer/RBSfinder"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="AAF84094.1"
    /db_xref="GI:9106274"
    /translation="MPITSPTRHAPKIFSPHNDIHQPSHSMHPNHQHTAQTNVSFAM
    HNPYLLKLTWLM"
    complement(2338..4227)
    /gene="XF1286"
    complement(2338..4227)
    /note="XF1286"
    /note="similar to SP131598 (percent identity: 64 %/query
    alignment_scores:

```

```

alignment coverage: 100.5 %/subject alignment coverage:
100.3 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit B"
/protein_id="AAF84095.1"
/db_xref="GI:9106275"
/translation="MNSRYNAADIEVLTLGDPVKRRPGMYTDTTRPNHLAQEVIDNSV
DEALAGYANRIEVTLYKDGCEVADNGRMPVDIHPEKIPGVLEILRLHAGGRFSN
RNYTSGGLHGVGVVNALSKVDLFIKREGSEHMFREFGAVSKLVGLTVSKKN
TTRLRFNPDQSKYFDTPKFNVRVRLHRAKAVLCPGLTWLLEPEATGEINTRHFKDG
LRDYLKGMETSELLPAELFICGLKADTEVDVMAVWYPELGTESVNLIPPTQHG
THVNGLSGLTDALREFCDRNLPRGKLAPEDVMDRVTVLSLUKMSDPQFSQGTKE
RUSKQAGFIETGVHDAPSLWNLQNDVIGEOIAQIATERASRLKTEKQVTRKKITH
GALPGKLADCIISLDLSRTFLVECDGAGGSAKARDKQFQAILPLRGKILNTWEVA
SGNVLASDEVHNLAIAGICPGKDLGLRYGVKVVILADSDGLHIALTLTALFLKH
FPALVAAGHVFMVAPLFRVDVKOVFALDDEKTMLEKIKREKIKGOVSVTRFKG
LGEMNPQLRESTIHPDTRRLVQLTMDONTQTCSLMDMLLAKKRASDRKQWLETGKNL
ASLDI"
4261..4701
/gene="XF1287"
4261..4701
/gene="XF1287"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84096.1"
/db_xref="GI:9106276"
/translation="MCCLLQIFARHSGSFSEGLAIPGRLGSGIPIEVALFRVTRGTSG
VATKLYLMLVPAVGVAVFSAQAQKASSTATEMQQPASPTDMPNTEGOG
MEHTEGHEKARKSKRHSHTVKTSTGATAPTAEENTSVNR"
4861..6525
/gene="XF1288"
4861..6525
/gene="XF1288"
/note="similar to SP108398 (percent identity: 65 %/query
alignment coverage: 97.1 %/subject alignment coverage:
98.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="CTP synthetase"
/protein_id="AAF84097.1"
/db_xref="GI:9106277"
/translation="MTPLIFVTGWSVSLGKIAAASLASILEARGLKVTWVKLDPYI
NVDPTWSPFOHGEVYVTDGAEATDLDLGHVFRVTRLRNNNSVTTGRVYQNVCKE
RRGYLGATQVVIPIHTEIRRCIDEATASFQVAVLVEIGTVGDIESLPLEAIRQVR
IERGARTMEHMLTLVPYIAAGELKTKPTQHSVKELRSIGIQPDVLLCKSEQVIDPS
ERRKIALFTNVSERAVICGPDIDVLYCMPLERLQGLDEIVDFQKLSGTFASLADLSE
WEDVVDAIKHPIDEVTIAVGVYVDQDAYKSGEALKHGLKQRKTVKNLKWVEAQL
EGSDMGALKDIDGLVPGFGDRGFKVLASRYARQRPVYFGICYGQMAAVVDYAR
HVAGLEGANSTENDRQSPHPVIALITEWRTTGTTEVERDEKSDMGDTMRCLQEORLK
AGTLVRELYGRDVVGRHRYEFNRYRTQLEDAGLVIAAKSMDDTLIELPREM
HPWFACQAHPEFLSTPDGHPFLFVGFKARAKKAGKLLREVCV"
6629..7516
/gene="XF1289"
6629..7516
/gene="XF1289"
/note="similar to SP17579 (percent identity: 44 %/query
alignment coverage: 94.6 %/subject alignment coverage:
98.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="2-dehydro-3-deoxyphosphoocetate aldolase"
/protein_id="AAF84098.1"
/db_xref="GI:9106278"
/translation="MRLSPFSLVLLVTPKQSGDVYKLCGFEVGLNQPLFLIAGPCVIES

```

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AE003962/rev ..

Align seg 1/1 to reverse of: AE003962 from: 1 to: 10986

143 LeuProAlaAsnThrGlnIleArg 150

1473 CTACACGACACACACAGATACGG 1450

seq_name: gb_ba:AE001354

seq_documentation_block:

LOCUS AE001354 11768 bp DNA linear BCT 30-OCT-2000

DEFINITION Chlamydia trachomatis section 81 of 87 of the complete genome.

ACCESSION AE001354 AE001273

VERSION AE001354.1 GI:3329280

KEYWORDS Chlamydia trachomatis.

SOURCE Chlamydia trachomatis.

ORGANISM Chlamydia trachomatis.

REFERENCE 1 (bases 1 to 11768)

AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,

Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,

Koonin,E.V. and Davis,R.W.

Genome sequence of an obligate intracellular pathogen of humans:

Chlamydia trachomatis

Science 282 (5389), 754-759 (1998)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 11768)

Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.

Comparative genomes of Chlamydia pneumoniae and C. trachomatis

Nat. Genet. 21 (4), 385-389 (1999)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

3 (bases 1 to 11768)

Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,

Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,

Koonin,E.V. and Davis,R.W.

Direct Submission

Submitted (20-MAY-1998) Program in Infectious Diseases, University

of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA

Location/Qualifiers

1..11768

/organism="Chlamydia trachomatis"

/strain="D/UW-3/CX"

/db_xref="taxon:813"

153..947

/gene="CT813"

153..947

/gene="CT813"

/transl_table=1

/product="hypothetical protein"

/protein_id="AAC68409.1"

/db_xref="GI:3329281"

/translation="MTPVNLCTNSNSINTFTKDIEMAKIQGRKDKPLAKTSWIA

LICVAVGLGAIIGGCSMASGLGIGAIVAVIGLCCLVSAICLQVEKQSWH

QKEFSWTEQSQFRIVNADMLKANRKLQSEVLEKSGWSDTAVHKEDVTKYEQVVE

EYAEIMELYETGVLTIEKINLQKRAKWLKEEAKMEQKLTITVDLEAAKQLEERV

TDLESEKQLEELDKATENDLDEMAEYEAEMEPEKEKKGKIPGRGSI"

complement(1031..1432)

/gene="CT814"

complement(1031..1432)

/gene="CT814"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

gene
 CDS
 protein_id="AAC68410.1"
 /db_xref="GI:3329282"
 /translation="MQOPPTLWSQTCRYSPRMALDKKRLTLTFVEKRMFSQPKKN
 KCLMLRGVLFQGLATLLSLFLPKSGMOIRKKLLRVKTSQTKKGRALLKSKHHT
 EFAEQTKLLAKNISKEIQDFTQSIIDSRD"
 complement(1351..1713)
 /gene="CT814.1"
 complement(1351..1713)
 /gene="CT814.1"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAC68411.1"
 /db_xref="GI:3329283"
 /translation="MTPYLIEALVACCLISLGMVLIFAFITICLLKLHKIMKLVSR
 SSLFEARWLAPLLIGKFKTLNWLKRYADRKMSQLADELEDNEDSENSCSRLLCG
 AKLAALCALCWLIRKND"
 1959..3335
 /gene="mrsA_2"
 /note="CT815"
 1959..3335
 /gene="mrsA_2"
 /codon_start=1
 /transl_table=1
 /product="Phosphoglucosyltransferase"
 /protein_id="AAC68412.1"
 /db_xref="GI:3329284"
 /translation="MTRDVRQLGDTGVRGRANPEPMTVETSVLLGKRAIGLVLEKHA
 GKRVVGVKDRLSGYPENALAGLTSMGIEITMLGPIPTPGVAFITIRADAGIM
 ISASHNPYKNGIKIFSDGDKFGKIGAEVEREAMVASKDFKDDHVCYCEPSSGN
 TGRYTEAKTEPKRTKLGRIVLDCAHGATYVAPSVFELDAEIVYCEPSSGN
 INAGCGALWPSTIQKAVIEHKADYGLDGDGRLIMVDEKGVHLOHMLNQAOLGSEOS
 KRRQALSDNRVVAVTMTNFGVRLYELSGIQVITISPGDRHVLQHMLNQAOLGSEOS
 GMMIFDNTTGDGIVSALQVLRIMIESESTLSLDTACIARSPQALINVPYTKVPLE
 SLANVQGVLYKEVKGSLGRILLRYSGTENICRMVMEGTTKTKHVDLSLAKTIVDVEA
 EIGARIS"
 3346..5166
 /gene="glms"
 /note="CT816"
 3346..5166
 /gene="glms"
 /codon_start=1
 /transl_table=1
 /product="Glucosamine-Fructose-6-P Amino transferase"
 /protein_id="AAC68413.1"
 /db_xref="GI:3329285"
 /translation="MCGIFGYLGERNAVPLVLEGLSKLEYRGYDSAGIATLVEGFLFV
 EKAVGVSQLCSAVSDIHSQAAIGHTRWATHGEPSEFRNAPHVDMDSALVHNGII
 ENFKLKELEQGVVFSDDTDEVIVOLFARRYKTRDLIOSFWSLKLQSGSFACA
 LMHODHPEVLICAAHESPLILGLGDEVFISDIIHAFKYSQCTQTLASGELAVLRIG
 KSTIYNFELARIQKEVRCIDHTEDSLDKGFDYMLKELYEQPEVFERILHLTCEN
 GFTESFLKGFSLDEICSLHIVAGSSYHAGYAKVYIESIASIPVYVETAGSEFRYRQ
 YIAESLAILISQSGETADTLAALNEFRKLKARVLGICNVRESALASRVDRCLFIEA
 GLEVGVAFTAKTAQTLILLGLRLANHRQVIAQEDLAQAIQGLKDLPLNLRFLDS
 SIHDMRCQIEETSFIFLGRFMYPCIMEAALUKKEIAYVEANAIAPAGEMKHGPTALI
 REGTPVIVYGDGSRVYTKTIGAIMEVKARKAYVIALAPESNRNDIAAASDEQIYIPDSH
 DLAAPIETIAGQINAYTMALQRGTEVDPRNLAKSVTVE"
 5269..6465
 /gene="tyrP_1"
 /note="CT817"
 5269..6465
 /gene="tyrP_1"
 /codon_start=1
 /transl_table=1
 /product="Tyrosine Transport"
 /protein_id="AAC68414.2"
 /db_xref="GI:6578114"
 /translation="MINKMLGALLVAGTTIGAGVLAVPVSTSEGFLPTLLLYIVSW
 FIYVAGSYCFLEVLTWTHSRKNVNMVMAEYTLGHKSKIIIMLVYLLFYSLLYVFC
 DGNILRMVWGCRSDTPIRHAMPVVFALFSPILMAKTSIIDOCNRFVFLGLGTF
 AMFCYGFPLMTDILLVRSANGTILKGFILFLATGCONVPTLHYMDKNVQVKKFA
 IVIGSSIPLVLIWIWAVLVGAVPISFLEQAKVEGTAIGALQTKSAFVYAGEFF
 GFFALISSFIGVSLGLKDFDFAQDMDEKRRKVEIFFLVFVFLVWVFPYGVILKCL

```

BCTGALGETIVLGVFPVLMWVGKRYGKRYGKRLPGGKSTLLVNSGLVLLNLVLLIA
6647..7840
/gene="tyrP_2"
/feature="Ct818"
6647..7840
/gene="tyrP_2"
/codon_start=1
/transl_table=11
/product="tyrosine transport"
/protein_id="AAC68415.1"
/db_xref="GI:3329287"
/translation="MCMNRKCVGILLVAGTVIGAGVLAVPIALTAVEGPPAVVLYVL
AWLYSLASGYGLVLTWCKGNQANCSMAEETLGVGRVLCVLYLTFYSLVAY
PCDGGNLSRVIGSFSPYPMRHMVPLLEFSLPAPLMMANTSVIDYCNRGFFVGLIF
VFGLLCVLPRTQGGELLRLASFSSLSNLSPIFFLAFQFONVPSLYHLDGNIRVK
RALLGSLFLPILITAEALVLTGTVPLVDLLKADGWTAGALGSLKNSAFYIAGE
LFFGFAVTSFGTALAKDFYIDFKWDARKRVSLFLVQVFLVWATFYPEIVLS
CLRYAGGAGGACIIIVPVMWLNMGYGRKRCFCGRILPGGKTVLLITGYTVNLAT
LYTTF"
7882..8598
/gene="ycca"
/feature="Ct819"
7882..8598
/gene="ycca"
/codon_start=1
/transl_table=11
/product="Transport Permease"
/protein_id="AAC68416.1"
/db_xref="GI:3329288"
/translation="MGLYDRDYAQDSRLPSTFSRRYVGMWTALVLTALTSGLVATG
AYRALFPMMWVINGFATLGVSYFYQAQIKLSVPVAVMGLFLAYSLRGMEFTLVPVYA
AQFGGVVWAAFGSAGIIFGLSAAGYAKNDLQJHRIIMLALVGLVLSLAFLIVS
LFTPMPLLYLLCYGLLIFVGLTVVDAQSIRVARSVDGHDLSKLSLIMALOMIC
NVLMIFWYLLQIFASDQR"
complement(8696..9550)
/gene="ftsY"
/feature="Ct820"
complement(8696..9550)
/gene="ftsY"
/codon_start=1
/transl_table=11
/product="Cell Division Protein FtsY"
/protein_id="AAC68417.1"
/db_xref="GI:3329289"
/translation="MEKFFGNLRSFLKTLSSDILEVAEVLVYEGDGPKLTEAFCE
ELRCKNPDERAVKELIRSELSKILSKLPQRPESLVRPFTSLVLTNGSGKTTVAKL
AHYLSQNKVLIVADTFRSAGMDQRCWATLNGCFISGPGDGAIAAFDGSAA
VADYDHYIDTSRGLHTHTNLKELOKATVCTNKAFFGAPGHETLMTIDATLGSNTLS
QVQLFHEAVPGLIETFEVGSAGSGSLFRIADKLPITRFVYGETIHDFFPFAIDR
FIDKLIDV"
9620..10780
/gene="sucC"
/feature="Ct821"
9620..10780
/gene="sucC"
/codon_start=1
/transl_table=11
/product="Succinyl-CoA Synthetase, Beta"
/protein_id="AAC68418.1"
/db_xref="GI:3329290"
/translation="MHLHXYQKDLLTAYQLDIPPHVATSVPEVETAIQAEQWAGV
VRAQVHAGRGKNGVVIHNSPQDLAAADKLLHMQFSNQTAGLSLPIKVLISPLV
EASEYLAIVIDRHRCPVIMLSKAGVDIEVAEKQPDQLLKMFLPSGKIYQYL
RRAKFMWDQIPADQGNRIIRQLQCFYKDAISLEINPLVTKDGLSLVLDKMTI
DNALYRHPQLADYDPSQENRDLVAKQLGSLYALDGTICLVNGAGLAMSTLID
KLYGGAANFLDVGSSASEKIQEALSLSVDPKSYVLEIFHFGIMQCAVVASGLVS
AMOGKETPTPTVRLCTNVDGKDMINAGIPCPEFVISMSEGELAVLSR"
10795..11670
/gene="sucD"

```

alignment_scores:
Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AE001354

Align seg 1/1 to: AE001354 from: 1 to: 11768

212 LeuGlyLeuYrAlaThrGlyAla 219

|||||

7993 CTAGGGTTATACGCAACAGGAGCG 8016

seq_name: gb_htg:AC110326

seq_documentation_block:

LOCUS AC110326

DEFINITION Rattus norvegicus clone CH230-296C10, *** SEQUENCING IN PROGRESS

*** 12 unordered pieces.

AC110326

AC110326.1 GI:18644766

VERSION HTG; HTGS_PHASE1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 18696)

REFERENCE

AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flaggs, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Jackson, L.E.,
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Joudah, S.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Kovar, C.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegid, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X.,
Rivers, M., Rojas, A., Rojokan, I., Roife, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, R.,
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (11-FEB-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOHV
Center clone name: CH230-296C10
----- Summary Statistics

Sequencing vector: Plasmid; M7789
Chemistry: Dye-terminator Big Dye; 98% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 18714 bases at least Q40
Consensus quality: 20119 bases at least Q30
Consensus quality: 20955 bases at least Q20
Estimated insert size: 13754; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1805: contig of 1805 bp in length
1806 1905: gap of unknown length
1906 3339: contig of 1434 bp in length
3340 3439: gap of unknown length
3440 5444: contig of 2005 bp in length
5445 5545: gap of unknown length
5546 7474: contig of 1929 bp in length
7475 7573: gap of unknown length
7574 8607: contig of 1034 bp in length
8608 10633: contig of 1926 bp in length
10634 10734: gap of unknown length
10735 12022: contig of 1289 bp in length
12023 12122: gap of unknown length
12123 13353: contig of 1231 bp in length
13354 13453: gap of unknown length
13454 14523: contig of 1070 bp in length
14524 14623: gap of unknown length
14624 16078: contig of 1455 bp in length
16079 16178: gap of unknown length
16179 17401: contig of 1223 bp in length
17402 17501: gap of unknown length
17502 18696: contig of 1195 bp in length.

```

FEATURES

Source

1. .18696
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-296C10"
BASE COUNT 4861 a 3968 c 3724 g 5041 t 1102 others
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AC110326/rev ..

Align seg 1/1 to reverse of: AC110326 from: 1 to: 18696

101 LeuLeuThrThrAlaArgSerTrp 108

|||||

4151 CTACTAACACAGCCAGAGCTGG 4128

seq_name: gb_in:CEK07A1

seq_documentation_block:

LOCUS CEK07A1 39086 bp DNA linear INV 24-JAN-2002

DEFINITION Caenorhabditis elegans cosmid K07A1, complete sequence.

ACCESSION Z81097

VERSION 281097.1 GI:1729626

KEYWORDS HTG; arginyl-tRNA protein transferase like; Chromatin assemble
factor 1 P55 subunit like; Deoxyuridine 5'-triphosphate
nucleotidylhydrolase; P58 protein like; Retinoblastoma-binding
protein RBAP46 like.

SOURCE

ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (sites)

REFERENCE

1 (sites)

none.

AUTHORS Genome sequence of the nematode C. elegans: a platform for

TITLE Investigating biology. The C. elegans Sequencing Consortium

JOURNAL Science 282 (5396), 2012-2018 (1998)

MEDLINE 9909613

REMARK The C.elegans Sequencing Consortium.

REFERENCE 2 (bases 1 to 39086)

AUTHORS Percy,C.M.

TITLE Direct Submission

JOURNAL Submitted (21-OCT-1996) Nematode Sequencing Project, Sanger

Institute, Hinxton, Cambridge CB10 1SA, England and Department of

Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:

jes@sanger.ac.uk or rwnematode.wustl.edu

On Dec 12, 1996 this sequence version replaced gi:1628014.

Coding sequences below are predicted from computer analysis, using

predictions from GeneFinder (P. Green, U. Washington), and other

available information.

Current sequence finishing criteria for the C. elegans genome

sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

This sequence is the entire insert of clone K07A1. The true right

end of clone K07A1 is at 16508 in this sequence. The start of this

sequence (1..105) overlaps with the end of sequence Z83225.

The end of this sequence (39087..39086) overlaps with the start of

sequence AL023833.

For a graphical representation of this sequence and its analysis

see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?>

name=K07A1

IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we only sequence

overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.

FEATURES

Location/Qualifiers

Source

gene

CDS

FEATURES

Source

gene

CDS

FEATURES

Source

gene

CDS

FEATURES

Source

gene

CDS

FEATURES

Source

gene

CDS

FEATURES

Source

gene

CDS

FEATURES

Source

gene

CDS

FEATURES

Source

gene

CDS


```

|||||
17484 TACCCTAATCTCGAATTAAATCAA 17461

seq_name: gb_pr:AL357372

seq_documentation_block:
LOCUS AL357372 43553 bp DNA linear PRI 20-JUL-2001
DEFINITION Human DNA sequence from clone RP11-86F19 on chromosome 10, complete
sequence.
ACCESSION AL357372
VERSION AL357372.12 GI:15020748
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 43553)
AUTHORS Wilson,S.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970346.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-86F19 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-86F19 it may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-399C16 is at 41554 in this
sequence. The true right end of clone RP11-177H22 is at 2000 in
this sequence.

FEATURES
Location/Qualifiers
1..43553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-86F19"
/clone_lib="RPCI-11.1"
2..213
/note="AluSg/x repeat: matches 144..282 of consensus"
repeat_region
213..273
/note="MLT1D repeat: matches 128..188 of consensus"
repeat_region
814..883
/note="35 copies 2 mer tt 65% conserved"
repeat_region
1079..1136
/note="29 copies 2 mer tt 72% conserved"
repeat_region
1235..1275
/note="MER4B repeat: matches 500..539 of consensus"
repeat_region
1385..1623
/note="158 copies 3 mer tct 71% conserved"

/note="L2 repeat: matches 1312..1543 of consensus"
1698..1944
/note="L2 repeat: matches 2040..2289 of consensus"
2015..2036
/note="11 copies 2 mer aa 100% conserved"
2100..2400
/note="AluSg repeat: matches 1..302 of consensus"
2838..3782
/note="L1M2 repeat: matches 5377..6298 of consensus"
3783..4088
/note="AluSx repeat: matches 2..307 of consensus"
4089..6514
/note="L1M2 repeat: matches 3010..5377 of consensus"
6515..6801
/note="AluSx repeat: matches 3..286 of consensus"
6802..6852
/note="L1M2 repeat: matches 2962..3010 of consensus"
6853..7139
/note="AluSc repeat: matches 1..287 of consensus"
7140..8465
/note="L1M2 repeat: matches 1669..2962 of consensus"
8466..8758
/note="AluSg repeat: matches 1..292 of consensus"
8759..8924
/note="L1M2 repeat: matches 1508..1669 of consensus"
8925..9223
/note="AluSx repeat: matches 1..312 of consensus"
9224..10543
/note="L1M2 repeat: matches 18..1508 of consensus"
10640..11114
/note="L1M2 repeat: matches 677..191 of consensus"
11200..11249
/note="L1M2 repeat: matches 151..200 of consensus"
11278..11373
/note="MIR repeat: matches 89..187 of consensus"
11793..11947
/note="L2 repeat: matches 1404..1573 of consensus"
11988..12104
/note="L2 repeat: matches 1825..1931 of consensus"
12105..12405
/note="AluY repeat: matches 1..302 of consensus"
12406..12774
/note="L2 repeat: matches 1931..2351 of consensus"
13807..14032
/note="MIR repeat: matches 20..257 of consensus"
14146..14445
/note="AluSx repeat: matches 1..307 of consensus"
14648..14957
/note="AluJb repeat: matches 1..312 of consensus"
15479..15615
/note="L2 repeat: matches 1663..1809 of consensus"
15650..15767
/note="MIR repeat: matches 49..178 of consensus"
16056..16346
/note="AluSx repeat: matches 1..292 of consensus"
16356..16683
/note="AluJo repeat: matches 1..310 of consensus"
16707..16958
/note="L2 repeat: matches 2466..2710 of consensus"
17120..17340
/note="AluJb repeat: matches 87..304 of consensus"
17544..17751
/note="MER20 repeat: matches 1..218 of consensus"
19449..19746
/note="AluSx repeat: matches 1..299 of consensus"
21326..21626
/note="AluSx repeat: matches 1..296 of consensus"
21890..22192
/note="AluSp repeat: matches 1..307 of consensus"
22395..22889
/note="11 copies 45 mer 71% conserved"
22398..22871
/note="158 copies 3 mer tct 71% conserved"

```

```

repeat_region 22400..22825
misc_feature /note="213 copies 2 mer cc 54% conserved"
22457..22542
/note="Sequence from clone PCR only"
repeat_region 22892..23059
/note="AluSg repeat: matches 1..168 of consensus"
repeat_region 23625..23934
/note="AluSg repeat: matches 3..310 of consensus"
24541..24687
/note="MIR repeat: matches 47..190 of consensus"
25743..25899
/note="MIR repeat: matches 79..243 of consensus"
26924..26973
/note="25 copies 2 mer at 76% conserved"
repeat_region 27039..27156
/note="L1MD3 repeat: matches 7625..7737 of consensus"
repeat_region 30352..30652
/note="AluSg repeat: matches 1..302 of consensus"
30668..30965
/note="AluY repeat: matches 1..297 of consensus"
31412..31543
/note="L1MD7 repeat: matches 6156..6289 of consensus"
31597..31624
/note="14 copies 2 mer tt 96% conserved"
repeat_region 33661..33781
/note="FLAM_C repeat: matches 1..121 of consensus"
34496..34693
/note="MIR repeat: matches 17..216 of consensus"
34897..35244
/note="AluY repeat: matches 1..310 of consensus"
36106..36324
/note="MIR repeat: matches 3..232 of consensus"
38610..38699
/note="2 copies 45 mer 96% conserved"
repeat_region 40346..40839
/note="L2 repeat: matches 2178..2713 of consensus"
40993..41212
/note="L2 repeat: matches 1990..2190 of consensus"
41213..41495
/note="AluSx repeat: matches 1..287 of consensus"
41496..41983
/note="L2 repeat: matches 1392..1990 of consensus"
41996..42088
/note="AluY repeat: matches 136..228 of consensus"
42144..42601
/note="L2 repeat: matches 732..1323 of consensus"
42903..43194
/note="AluSg repeat: matches 1..292 of consensus"
BASE COUNT 11508 a 8161 c 8761 g 15123 t
ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
us-09-674-779-2 x AL357372/rev ..

Align seg 1/1 to reverse of: AL357372 from: 1 to: 43553

168 SerLysHisLeuThrAsnSerAla 175
|||||
11182 AGTAACATCTGACAAACAGTGCC 11159

seq_name: gb_htg:AC098224

seq_documentation_block:
LOCUS AC098224
DEFINITION Rattus norvegicus clone CH230-34C16, *** SEQUENCING IN PROGRESS
***, 30 unordered pieces.
ACCESSION AC098224

```

```

VERSION AC098224.3 GI:17974394
KEYWORDS HTGS_PHASE1
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 44428)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Drapet,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsy,F., Howard,S., Huber,J., Huiyik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shostari,N.,
Sisson,L., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stang,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczky,R., Woodden,S.,
Worley,K., Wu,C., Wu,Y., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 44428)
Worley,K.C.
Direct Submission
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17062625.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GIVE
Center clone name: CH230-34C16
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 36210 bases at least Q40
Consensus quality: 39629 bases at least Q30
Consensus quality: 41567 bases at least Q20
Estimated insert size: 11686; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation

```

```

TITLE JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 30 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1856: contig of 1856 bp in length
1857 1956: gap of unknown length
1957 4072: contig of 2116 bp in length
4073 4172: gap of unknown length
4173 5702: contig of 1530 bp in length
5703 5802: gap of unknown length
5803 7463: contig of 1661 bp in length
7464 7563: gap of unknown length
7564 9576: contig of 2013 bp in length
9577 9676: gap of unknown length
9677 11737: contig of 2061 bp in length
11738 11837: gap of unknown length
11838 13410: contig of 1573 bp in length
13411 13510: gap of unknown length
13511 14639: contig of 1129 bp in length
14640 16113: contig of 1374 bp in length
16114 16213: gap of unknown length
16214 17474: contig of 1261 bp in length
17475 17574: gap of unknown length
17575 18813: contig of 1239 bp in length
18814 18913: gap of unknown length
18914 20077: contig of 1164 bp in length
20078 20177: gap of unknown length
20178 21581: contig of 1404 bp in length
21582 21681: gap of unknown length
21682 22975: contig of 1294 bp in length
22976 23075: gap of unknown length
23076 24104: contig of 1029 bp in length
24105 24204: gap of unknown length
24205 25556: contig of 1252 bp in length
25557 26776: gap of unknown length
26777 26876: contig of 1220 bp in length
26877 28191: contig of 1315 bp in length
28192 28291: gap of unknown length
28292 29372: contig of 1081 bp in length
29373 29472: gap of unknown length
29473 30582: contig of 1110 bp in length
30583 32269: contig of 1587 bp in length
32270 32369: gap of unknown length
32370 33466: contig of 1097 bp in length
33467 35103: contig of 1537 bp in length
35104 35203: gap of unknown length
35204 36237: contig of 1034 bp in length
36238 36337: gap of unknown length
36338 37837: contig of 1500 bp in length
37838 37937: gap of unknown length
37938 38961: contig of 1024 bp in length
38962 39061: gap of unknown length
39062 40129: contig of 1068 bp in length
40130 40229: gap of unknown length
40230 41866: contig of 1637 bp in length
41867 41966: gap of unknown length
41967 43065: contig of 1099 bp in length
43066 43155: gap of unknown length
43156 44428: contig of 1263 bp in length.

FEATURES
source

Location/Qualifiers
1. .44428
/organism="Rattus norvegicus"

/db_xref="taxon:10116"
/clone="CH230-34C16"
BASE COUNT 8998 a 11665 c 11518 g 9322 t 2925 others
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x AC098224/rev ..
Align seg 1/1 to reverse of: AC098224 from: 1 to: 44428
138 LysSerArgGlyIleLeuProAla 145
|||||
3048 AAAAGTAGGGGTATCTTGCCAGCA 3025
seq_name: gb_htg:AC014969
seq_documentation_block:
LOCUS AC014969 46275 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014969
VERSION AC014969.1 GI:6436366
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 46275)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210723 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .46275
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 12910 a 10354 c 10692 g 12319 t
ORIGIN
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x AC014969/rev ..
Align seg 1/1 to reverse of: AC014969 from: 1 to: 46275
8 PheIleThrThrLeuIleSerSer 15
|||||
16423 TTCAATTACTACGCTTATAGCTCT 16400

OM of: US-09-674-779-2 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 16, 2002 7:56 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO/spool/US09674779/runat_12092002_124144_26892/app_query.fasta_1.310
-DB-Issued_Patents_NA -QWTFastap -SUFFIX=oli.rni -GAPOP=4.500
-GAPEXT=0.050 -MINWATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -GAPEXT=60.000
-DELOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -GAPEXT=60.000
-TRANS=human40.cdi -LIST=45 -DALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -CGN1_1_47 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WGN1_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-2
Query length: 250
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 45.910000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-365-486A-14 +		8.00	152.83	0.7978	49 ! Patent
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-080-285-19 +		8.00	127.99	19.30	1242 ! Pate
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-880-342-14 +		8.00	127.86	19.61	1262 ! Pate
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-7 -		8.00	124.30	30.97	2007 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-743-637B-268 +		7.00	113.11	130.10	692 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-998-416-378 +		7.00	112.00	149.90	799 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-815-688A-4 -		7.00	108.11	247.14	1128 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-507-431-3 +		7.00	108.11	247.14	1327 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-902-655A-3 +		7.00	108.11	247.14	1327 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-116-622-3 +		7.00	108.11	247.14	1327 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-219-277-3 +		7.00	108.11	247.14	1327 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-599-661-3 +		7.00	108.11	247.14	1327 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-000-630C-1 +		7.00	106.16	317.31	1710 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-862-730C-1 +		7.00	106.16	317.31	1710 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-202-056-4 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-076-093A-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-701-265-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-284-586-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-805-478-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-802-627A-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-238-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-228-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-104-296-3 +		7.00	106.04	322.25	1737 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-148-680-1 +		7.00	104.98	369.00	1993 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-522-6 -		7.00	104.33	400.92	2168 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-131-648-4 -		7.00	103.91	423.14	2290 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-176-620A-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-463-862-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-461-985-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-458-887-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-012C-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-888-818C-5 +		7.00	100.29	673.71	3671 !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-5459251-1 +		7.00	98.19	882.01	4825 !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-030-096-7 -		7.00	97.95	909.38	4977 !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-485A-19 +		7.00	97.78	929.01	5086 !

/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-365-486A-14 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-080-285-19 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-880-342-14 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-7 - 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-743-637B-268 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-998-416-378 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-815-688A-4 - 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-507-431-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-902-655A-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-116-622-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-219-277-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-599-661-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-000-630C-1 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-862-730C-1 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-202-056-4 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-076-093A-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-701-265-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-284-586-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-805-478-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-802-627A-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-238-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-228-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-104-296-3 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-148-680-1 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-522-6 - 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-131-648-4 - 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-176-620A-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-463-862-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-461-985-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-458-887-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-012C-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-888-818C-5 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-5459251-1 + 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-030-096-7 - 7.00 97.78 929.01 5086
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-485A-19 + 7.00 97.78 929.01 5086

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-19

seq_documentation_block:

; Patent No. 5240845
; APPLICANT: FUJII, SETSURO; TAKADA, KAORUKO; . KATANO, TAMIKI;
; MAJIMA, EIJI; OGINO, KOICHI; ONO, KENJI; SAKATA, YASUYO; UENOYAMA,
; TSUTOMU
; TITLE OF INVENTION: MUTATED STREPTOKINASE PROTEINS
; NUMBER OF SEQUENCES: 65
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/549,049
; FILING DATE: 06-JUL-1990
; SEQ ID NO:19:
; LENGTH: 49
; 5240845-19

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x 5240845-19/rev ..

Align seg 1/1 to reverse of: 5240845-19 from: 1 to: 49

192 GluLeuGlnAsnArgLeuCysGln 199
|||||
26 GAGTTACAGAACCGCTTTGTCAG 3

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-2

seq_documentation_block:

; Patent No. 5240845
; APPLICANT: FUJII, SETSURO; TAKADA, KAORUKO; . KATANO, TAMIKI;
; MAJIMA, EIJI; OGINO, KOICHI; ONO, KENJI; SAKATA, YASUYO; UENOYAMA,
; TSUTOMU
; TITLE OF INVENTION: MUTATED STREPTOKINASE PROTEINS
; NUMBER OF SEQUENCES: 65
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/549,049
; FILING DATE: 06-JUL-1990
; SEQ ID NO:2:
; LENGTH: 1242
; 5240845-2

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x 5240845-2/rev ..

Align seg 1/1 to reverse of: 5240845-2 from: 1 to: 1242

192 GluLeuGlnAsnArgLeuCysGln 199
|||||
379 GAGTTACAGAACCGCTTTGTCAG 356

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5240845-3

seq_documentation_block:

```
; Patent No. 5240845
; APPLICANT: FUJII, SETSURO; TAKADA, KAORUKO; . KATANO, TAMIKI;
; MAJIMA, EIJI; OGINO, KOICHI; ONO, KENJI; SAKATA, YASUYO; UENOYAMA,
; TSUTOMU
; TITLE OF INVENTION: MUTATED STREPTOKINASE PROTEINS
; NUMBER OF SEQUENCES: 65
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/549,049
; FILING DATE: 06-JUL-1990
; SEQ ID NO: 3:
; LENGTH: 1262
5240845-3

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x 5240845-3/rev ..
Align seg 1/1 to reverse of: 5240845-3 from: 1 to: 1262

192 GluLeuGlnAsnArgLeuCysGln 199
|||||
393 GAGTTACAGAACCGCTTTGTCAG 370

seq_name: /cgn2.6/ptodata/2/ina/6B_COMB.seq:US-08-552-369-7

seq_documentation_block:
; Sequence 7, Application US/08552369
; Patent No. 6241989
; GENERAL INFORMATION:
; APPLICANT: Scott, Fred W.
; APPLICANT: Ngichabe, Christopher K.
; APPLICANT: Hu, Liandbiao
; TITLE OF INVENTION: Recombinant Multivalent Viral Vaccine
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear
; STREET: 1800 One Mt Plaza
; CITY: Buffalo
; STATE: New York
; COUNTRY: United States
; ZIP: 14203-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/552,369
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/190,789
; FILING DATE: 01/27/1994
; APPLICATION NUMBER: 07/726,609
; FILING DATE: 07/09/1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Nelson, M. Bud
; REGISTRATION NUMBER: 35,300
; REFERENCE/DOCKET NUMBER: 18617.0016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 856-4000
; TELEFAX: (716) 849-0349
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2007 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
```

```
; HYPOTHETICAL: yes
; ORIGINAL SOURCE:
; ORGANISM: feline calicivirus
; FEATURE:
; LOCATION: capsid protein gene region
; US-08-552-369-7

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-552-369-7/rev ..
Align seg 1/1 to reverse of: US-08-552-369-7 from: 1 to: 2007

192 GluLeuGlnAsnArgLeuCysGln 199
|||||
1090 GAATTACAAAATCGTGTATGTCAG 1067

seq_name: /cgn2.6/ptodata/2/ina/5B_COMB.seq:US-08-743-637B-268

seq_documentation_block:
; Sequence 268, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C. 35,433
; REGISTRATION NUMBER: 850586.90012
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 268:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus mutans
; US-08-743-637B-268
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-674-779-2 x US-08-743-637B-268  ..
  Align seg 1/1  to: US-08-743-637B-268  from: 1  to: 692

      58 SerHisValGlyLeuGlnAla 64
      |||||
289 AGTCATGTTGGCTTACACCA 309

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-998-416-378

seq_documentation_block:
; Sequence 378, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtie, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 378:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 799 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1287RP
US-08-998-416-378

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
```

```
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-674-779-2 x US-08-998-416-378  ..
  Align seg 1/1  to: US-08-998-416-378  from: 1  to: 799

      12 LeuIleSerSerMetLeuVal 18
      |||||
490 TTAATATCTTCATGCTTGTT 510

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-815-688A-4

seq_documentation_block:
; Sequence 4, Application US/08815688A
; Patent No. 5786195
; GENERAL INFORMATION:
; APPLICANT: Xu, Shuang-yong
; APPLICANT: Xiao, Jian-ping
; TITLE OF INVENTION: METHOD FOR CLONING AND
; TITLE OF INVENTION: PRODUCING THE BSSHII RESTRICTION ENDONUCLEASE
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/815,688A
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-927-5054
; TELEFAX: 508-927-1705
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding sequence
; LOCATION: 1..1125
; OTHER INFORMATION:
US-08-815-688A-4

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-674-779-2 x US-08-815-688A-4/rev  ..
  Align seg 1/1  to reverse of: US-08-815-688A-4  from: 1  to: 1128
```

60 valglyLeuGlnAlaHisPhe 66
|||||
676 GTAGGTTGCGAGGCCCATTTTC 696
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-700-546-1
seq_documentation_block:
; Sequence 1, Application US/08700546
; Patent No. 5874274
; GENERAL INFORMATION:
; APPLICANT: Jacobsen, Tina Sejersgard
; APPLICANT: Heldt-Hansen, Hans Peter
; APPLICANT: Kofod, Lene Venke
; APPLICANT: Bagger, Christian
; APPLICANT: Mullertz, Anette
; TITLE OF INVENTION: Process For Processing Plant Material
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5874274o No. 5874274disk of No. 5874274th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/700,546
; FILING DATE: 23-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4141.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4..1221, 1225..1314, 1318..1326)
US-08-700-546-1
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x US-08-700-546-1 ..
Align seg 1/1 to: US-08-700-546-1 from: 1 to: 1327
60 valglyLeuGlnAlaHisPhe 66
|||||
676 GTAGGTTGCGAGGCCCATTTTC 696
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-902-655A-3
seq_documentation_block:
; Sequence 3, Application US/08902655A
; Patent No. 5885819
; GENERAL INFORMATION:

18 ValAlaCysSerAlaProfile 24
|||||
536 GTAGCTTGTTCAGCACCAATA 516
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-507-431-3
seq_documentation_block:
; Sequence 3, Application US/08507431
; Patent No. 5693518
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Heldt-Hansen, Hans P.
; APPLICANT: Dalboge, Henrik
; APPLICANT: Andersen, Lene N.
; APPLICANT: Si, Joan O.
; APPLICANT: Jacobson, Tina
; APPLICANT: Munk, Niels
; APPLICANT: Mullertz, Anette
; TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5693518o No. 5693518disk of No. 5693518th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/507,431
; FILING DATE: 15-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/002,800
; FILING DATE: 25-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 3954.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4..1221, 1225..1314, 1318..1326)
US-08-507-431-3
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x US-08-507-431-3 ..
Align seg 1/1 to: US-08-507-431-3 from: 1 to: 1327


```

; APPLICANT: Kofod, Lene V.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Heldt-Hansen, Hans P.
; APPLICANT: Dalboge, Henrik
; APPLICANT: Andersen, Lene N.
; APPLICANT: Si, Joan Q.
; APPLICANT: Jacobson, Tina
; APPLICANT: Mullertz, Anette
; APPLICANT: Munk, Niels
; APPLICANT: Mullertz, Anette
; TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
; TITLE OF INVENTION: ASPERGILLUS ACULEATUS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 58858190 No. 58858190disk of No. 5885819th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902.655A
; FILING DATE: 30-July-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl T.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3954.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(4..1221, 1225..1314, 1318..1326)
; US-08-902-655A-3

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
  US-09-674-779-2 x US-08-902-655A-3  ..
  Align seg 1/1 to: US-08-902-655A-3  from: 1 to: 1327

```

```

60 ValGlyLeuGlnAlaHisPhe 66
|||||
676 GTAGGGTTGCAGGCCCATTC 696

```

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-116-622-3

```

```

seq_documentation_block:
; Sequence 3, Application US/09116622
; Patent No. 6080567
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Heldt-Hansen, Hans P.

```

```

; APPLICANT: Dalboge, Henrik
; APPLICANT: Andersen, Lene N.
; APPLICANT: Si, Joan Q.
; APPLICANT: Jacobson, Tina
; APPLICANT: Munk, Niels
; APPLICANT: Mullertz, Anette
; TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
; TITLE OF INVENTION: ASPERGILLUS ACULEATUS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6080567o No. 6080567disk of No. 6080567th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/116.622
; FILING DATE: 16-July-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Agris, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3954.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(4..1221, 1225..1314, 1318..1326)
; US-09-116-622-3

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
  US-09-674-779-2 x US-09-116-622-3  ..
  Align seg 1/1 to: US-09-116-622-3  from: 1 to: 1327

```

```

60 ValGlyLeuGlnAlaHisPhe 66
|||||
676 GTAGGGTTGCAGGCCCATTC 696

```

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-219-277-3

```

```

seq_documentation_block:
; Sequence 3, Application US/09219277
; Patent No. 6197564
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan
; APPLICANT: Heldt-Hansen, Hans P.
; APPLICANT: Dalboge, Henrik
; APPLICANT: Andersen, Lene N.
; APPLICANT: Si, Joan Q.
; APPLICANT: Jacobson, Tina

```

Tue Sep 17 07:27:53 2002

```
;
; APPLICANT: Munk, Niels
; APPLICANT: Mullertz, Anette
; TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 61975640 No. 6197564disk of No. 6197564th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/219,277
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/116,622
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3954.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4..1221, 1225..1314, 1318..1326)
;
US-09-219-277-3
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-674-779-2 x US-09-219-277-3 ..
```

```
Align seg 1/1 to: US-09-219-277-3 from: 1 to: 1327
```

```
60 valglyLeuGlnAlaHisPhe 66
|||||
676 GTAGGTTGCAGGCCCATTC 696
```

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-599-661-3
```

```
seq_documentation_block:
; Sequence 3, Application US/09599661
; Patent No. 6228630
```

```
; GENERAL INFORMATION:
; APPLICANT: Kofod, Lene V.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Christgau, Stephan P.
; APPLICANT: Heldt-Hansen, Hans P.
; APPLICANT: Dalboge, Henrik
; APPLICANT: Andersen, Lene N.
; APPLICANT: Si, Joan Q.
; APPLICANT: Jacobson, Tina
; APPLICANT: Munk, Niels
```

```
;
; APPLICANT: Mullertz, Anette
; TITLE OF INVENTION: ENZYMES WITH XYLANASE ACTIVITY FROM
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 62286300 No. 6228630disk of No. 6228630th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/599,661
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/116,622
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3954.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4..1221, 1225..1314, 1318..1326)
;
US-09-599-661-3
```

```
alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-674-779-2 x US-09-599-661-3 ..
```

```
Align seg 1/1 to: US-09-599-661-3 from: 1 to: 1327
```

```
60 ValglyLeuGlnAlaHisPhe 66
|||||
676 GTAGGTTGCAGGCCCATTC 696
```

```
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-000-630C-1
```

```
seq_documentation_block:
; Sequence 1, Application US/09000630C
; Patent No. 6018029
```

```
; GENERAL INFORMATION:
; APPLICANT: Fuller, Gerald M.
; APPLICANT: Fuentes, Nelson L.
; TITLE OF INVENTION: DNA Encoding Canine Interleukin-1 Receptor
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Douglas C Murdock/ Bradley, Arant, Rose & White
; STREET: 2001 Park Place, Suite 1400
; CITY: Birmingham
; STATE: Alabama
```

```

/ COUNTRY: USA
/ ZIP: 35203-2736
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch,
/ OPERATING SYSTEM: Microsoft Windows
/ SOFTWARE: WordPerfect 6.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/000,630C
/ FILING DATE:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/862,730
/ FILING DATE:
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1710 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA to mRNA
/ HYPOTHETICAL: N
/ ANTI-SENSE: N
/ ORIGINAL SOURCE:
/ ORGANISM: Canis familiaris
/ CELL TYPE: canine peripheral blood macrophage
/ CELL LINE: primary monocytes
/ IMMEDIATE SOURCE:
/ LIBRARY: lambda gt11 cDNA
/ CLONE: Canine IL-1ra
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1 to 1710
/ OTHER INFORMATION:
/ FEATURE:
/ NAME/KEY: open reading frame
/ LOCATION: 60 to 587
/ OTHER INFORMATION:
/ US-09-000-630C-1

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-09-000-630C-1 ..
Align seg 1/1 to: US-09-000-630C-1 from: 1 to: 1710

31 SerProleLysThrProSer 37
|||||
1358 TCTCCATCAAACTCCAGC 1378

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-862-730C-1

seq_documentation_block:
; Sequence 1, Application US/08862730C
; Patent No. 6063600
; GENERAL INFORMATION:
; APPLICANT: Fuller, Gerald M
; APPLICANT: Fuentes, Nelson L
; TITLE OF INVENTION: DNA Encoding Canine Interleukin-1 Receptor
; TITLE OF INVENTION: Antagonist
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Douglas C Muddock/ Bradley, Arant, Rose & White
; STREET: 2001 Park Place, Suite 1400
; CITY: Birmingham
; STATE: Alabama
; COUNTRY: USA
; ZIP: 35203-2736
; COMPUTER READABLE FORM:

```

```

/ MEDIUM TYPE: Diskette, 3.50 inch,
/ COMPUTER: IBM compatible
/ OPERATING SYSTEM: Microsoft Windows
/ SOFTWARE: WordPerfect 6.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/862,730C
/ FILING DATE: 5/23/97
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1710 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA to mRNA
/ HYPOTHETICAL: N
/ ANTI-SENSE: N
/ ORIGINAL SOURCE:
/ ORGANISM: Canis familiaris
/ CELL TYPE: canine peripheral blood macrophage
/ CELL LINE: primary monocytes
/ IMMEDIATE SOURCE:
/ LIBRARY: lambda gt11 cDNA
/ CLONE: Canine IL-1ra
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1 to 1710
/ OTHER INFORMATION:
/ FEATURE:
/ NAME/KEY: open reading frame
/ LOCATION: 60 to 587
/ OTHER INFORMATION:
/ US-08-862-730C-1

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-862-730C-1 ..
Align seg 1/1 to: US-08-862-730C-1 from: 1 to: 1710

31 SerProleLysThrProSer 37
|||||
1358 TCTCCATCAAACTCCAGC 1378

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-202-056-4

seq_documentation_block:
; Sequence 4, Application US/08202056
; Patent No. 5440021
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Hebert, Caroline
; APPLICANT: Kim, Kyung Jin
; APPLICANT: Lee, James
; TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/202,056
; FILING DATE: 25-FEB-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-202-056-4

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

alignment_block:

```
US-09-674-779-2 x US-08-202-056-4 ..
```

```
Align seg 1/1 to: US-08-202-056-4 from: 1 to: 1737
```

```

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTCTGGCGGCGGCG 27

```

```
seq_name: /cgn2,6/ptodata/2/ina/5A_COMB.seq:US-08-076-093A-3
```

```

seq_documentation_block:
; Sequence 3, Application US/08076093A
; Patent No. 5543503
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B

```

```

; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-076-093A-3

```

```

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```
alignment_block:
US-09-674-779-2 x US-08-076-093A-3 ..
```

```
Align seg 1/1 to: US-08-076-093A-3 from: 1 to: 1737
```

```

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTCTGGCGGCGGCG 27

```

```
seq_name: /cgn2,6/ptodata/2/ina/5A_COMB.seq:US-08-701-265-3
```

```

seq_documentation_block:
; Sequence 3, Application US/08701265
; Patent No. 5776457
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,265
; FILING DATE: 22-AUG-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-Jun-1993
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168

```

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-701-265-3

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-701-265-3 ..
Align seg 1/1 to: US-08-701-265-3 from: 1 to: 1737

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-284-586-3

seq_documentation_block:
; Sequence 3, Application US/08284586
; Patent No. 5840856
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: Antibodies to Human PF4A Receptors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,586
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,093A
; FILING DATE: 11-Jun-1993
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/677211
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: 706P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear

```

```

US-08-284-586-3

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-284-586-3 ..
Align seg 1/1 to: US-08-284-586-3 from: 1 to: 1737

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGGCGGCG 27

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-805-478-3

seq_documentation_block:
; Sequence 3, Application US/08805478
; Patent No. 5874543
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: Jin Kim, K.
; TITLE OF INVENTION: ANTIBODIES TO PF4A RECEPTOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,478
; FILING DATE: 25-Feb-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284586
; FILING DATE: 10-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: P0706P2P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-805-478-3

alignment_scores:
  Quality: 7.00      Length: 7

```

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x US-08-805-478-3 ..
Align seg 1/1 to: US-08-805-478-3 from: 1 to: 1737

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGCGCGCG 27
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-802-627A-3
seq_documentation_block:
; Sequence 3, Application US/08802627A
; Patent No. 5892017
; GENERAL INFORMATION:
; APPLICANT: Lee, James
; TITLE OF INVENTION: NUCLEIC ACID ENCODING PF4A RECEPTOR
; NUMBER OF SEQUENCES: 6
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,627A
; FILING DATE: 19-Feb-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284586
; FILING DATE: 10-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: P0706P2P1D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-802-627A-3

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x US-08-802-627A-3 ..
Align seg 1/1 to: US-08-802-627A-3 from: 1 to: 1737

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGCGCGCG 27
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-238-3
seq_documentation_block:
; Sequence 3, Application US/08801238
; Patent No. 591896
; GENERAL INFORMATION:
; APPLICANT: Lee, James
; TITLE OF INVENTION: PF4A RECEPTOR
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,238
; FILING DATE: 19-Feb-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284586
; FILING DATE: 10-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: P0706P2P1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
US-08-801-238-3

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x US-08-801-238-3 ..
Align seg 1/1 to: US-08-801-238-3 from: 1 to: 1737

160 GlnCysAlaGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGCGCGCG 27
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-801-228-3

```

seq_documentation_block:
; Sequence 3, Application US/08801228
; Patent No. 5922541
; GENERAL INFORMATION:
; APPLICANT: Lee, James
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: METHODS FOR DETECTION AND AMPLIFICATION OF
; TITLE OF INVENTION: PF4A RECEPTOR NUCLEIC ACID
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,228
; FILING DATE: 19-Feb-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284586
; FILING DATE: 10-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: P0706P2P1D3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-801-228-3

```

```

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-674-779-2 x US-08-801-228-3 ..

```

```

Align seg 1/1 to: US-08-801-228-3 from: 1 to: 1737

```

```

160 GlnCysAlaGlyGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGGCGGCG 27

```

```

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-104-296-3

```

```

seq_documentation_block:
; Sequence 3, Application US/09104296
; Patent No. 6087475
; GENERAL INFORMATION:
; APPLICANT: Lee, James
; APPLICANT: Wood, William I.

```

```

; TITLE OF INVENTION: PF4A Receptors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,296
; FILING DATE: 24-June-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/701265
; FILING DATE: 22-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/664228
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076093
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/810782
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Love, Richard B.
; REGISTRATION NUMBER: 34,659
; REFERENCE/DOCKET NUMBER: P0706P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5530
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1737 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-09-104-296-3

```

```

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:
US-09-674-779-2 x US-09-104-296-3 ..

```

```

Align seg 1/1 to: US-09-104-296-3 from: 1 to: 1737

```

```

160 GlnCysAlaGlyGlyAlaAla 166
|||||
7 CAGTGTGCTGGCGGCGGCG 27

```

```

seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:PCT-US94-06380-2

```

```

seq_documentation_block:
; Sequence 2, Application PC/TUS9406380
; GENERAL INFORMATION:
; APPLICANT: Chuntharapai, Anan
; APPLICANT: Lee, James
; APPLICANT: Hebert, Caroline
; APPLICANT: K. Jin Kim
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: METHODS FOR TREATING INFLAMMATORY DISORDERS
; NUMBER OF SEQUENCES: 3

```


;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Genentech, Inc.
;; STREET: 460 Point San Bruno Blvd
;; CITY: South San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94080
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: patin (Genentech)
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US94/06380
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/076093
;; FILING DATE: 11-JUN-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Love, Richard B.
;; REGISTRATION NUMBER: 34,659
;; REFERENCE/DOCKET NUMBER: 706P2P1
;; TELEPHONE: 415/225-5530
;; TELEFAX: 415/952-9881
;; TELEX: 910/371-7168
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1737 bases
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; PCT-US94-06380-2
;;
;; alignment_scores:
;; Quality: 7.00 Length: 7
;; Ratio: 1.000 Gaps: 0
;; Percent Similarity: 100.000 Percent Identity: 100.000
;;
;; alignment_block:
;; US-09-674-779-2 x PCT-US94-06380-2 ..
;;
;; Align seg 1/1 to: PCT-US94-06380-2 from: 1 to: 1737
;;
160 GlnCysAlaGlyClyAlaAla 166
|||||
7 CAGTGTGCTGGCGGCGGCGG 27
;;
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-148-680-1
;;
seq_documentation_block:
;; Sequence 1, Application US/09148680
;; Patent No. 6255561
;; GENERAL INFORMATION:
;; APPLICANT: Kossmann, Jens
;; APPLICANT: Willmitzer, Lothar
;; APPLICANT: Emmermann, Michael
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES CODING FOR DEBRANCHING
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FISH & NEAVE
;; STREET: 1251 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10020-1104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;;
;; alignment_scores:
;; Quality: 7.00 Length: 7
;; Ratio: 1.000 Gaps: 0
;; Percent Similarity: 100.000 Percent Identity: 100.000
;;
;; alignment_block:
;; US-09-674-779-2 x US-09-148-680-1 ..
;;
;; Align seg 1/1 to: US-09-148-680-1 from: 1 to: 1993
;;
150 ArgSerValTyrArgAsnPro 156
|||||
1509 CCAAGTGTCTATAGAAATCCC 1529
;;
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-522-6
;;
seq_documentation_block:
;; Sequence 6, Application US/08749522
;; Patent No. 6096950
;; GENERAL INFORMATION:
;; APPLICANT: John, Maliyakal
;; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles & Brady
;; STREET: 411 East Wisconsin Avenue
;; CITY: Milwaukee
;; STATE: WI
;; COUNTRY: U.S.A.
;; ZIP: 53202-4497
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/749,522

;; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPA)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/148,680
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/EP97/01141
;; FILING DATE: 06-MAR-97
;; APPLICATION NUMBER: DE 196 08 918.2
;; FILING DATE: 07-MAR-96
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Haley Jr., James F.
;; REGISTRATION NUMBER: 27,794
;; REFERENCE/DOCKET NUMBER: GFB-7
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 596-9000
;; TELEFAX: (212) 596-9090
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1993 base pairs
;; TYPE: nucleotide
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Zea mays
;; TISSUE TYPE: Blattgewebe
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..1675
;; US-09-148-680-1
;;
;; alignment_scores:
;; Quality: 7.00 Length: 7
;; Ratio: 1.000 Gaps: 0
;; Percent Similarity: 100.000 Percent Identity: 100.000
;;
;; alignment_block:
;; US-09-674-779-2 x US-09-148-680-1 ..
;;
;; Align seg 1/1 to: US-09-148-680-1 from: 1 to: 1993
;;
150 ArgSerValTyrArgAsnPro 156
|||||
1509 CCAAGTGTCTATAGAAATCCC 1529
;;
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-749-522-6
;;
seq_documentation_block:
;; Sequence 6, Application US/08749522
;; Patent No. 6096950
;; GENERAL INFORMATION:
;; APPLICANT: John, Maliyakal
;; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles & Brady
;; STREET: 411 East Wisconsin Avenue
;; CITY: Milwaukee
;; STATE: WI
;; COUNTRY: U.S.A.
;; ZIP: 53202-4497
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/749,522

```

; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C. 35,433
; REGISTRATION NUMBER: 670513.90244
; REFERENCE/DOCKET NUMBER: 670513.90244
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-749-522-6

alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-749-522-6/rev ..

Align seg 1/1 to reverse of: US-08-749-522-6 from: 1 to: 2168

```

```

136 AspLeuLysSerArgGlyIle 142
|||||
338 GATTAAAGATGACAGGATT 318

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-131-648-4

```

```

seq_documentation_block:
; Sequence 4, Application US/09131648
; Patent No. 6168920
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
; FILE REFERENCE: PF-0576 US
; CURRENT APPLICATION NUMBER: US/09/131.648
; CURRENT FILING DATE: 1998-08-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 2290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2687731
US-09-131-648-4

```

```

alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-09-131-648-4 ..

Align seg 1/1 to: US-09-131-648-4 from: 1 to: 2290

```

```

143 LeuProAlaAsnThrGluIle 149
|||||
261 TTGCAGCTAACACACAGATT 281

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-176-620A-5

seq_documentation_block:
; Sequence 5, Application US/08176620A
; Patent No. 5595904
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G.
; APPLICANT: Cobb, Melanie H.
; APPLICANT: Yancopoulos, George D.
; APPLICANT: Nye, Steven
; APPLICANT: Panayotatos, Nikos
; TITLE OF INVENTION: A Family of Map2 Protein Kinases
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,620A
; FILING DATE: 03-JAN-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6526-123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 303..2018
US-08-176-620A-5

```

```

alignment_scores:
    Quality: 7.00      Length: 7
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

```

```

alignment_block:
US-09-674-779-2 x US-08-176-620A-5 ..

```

```

Align seg 1/1 to: US-08-176-620A-5 from: 1 to: 3671

```

```

8 PheIleThrThrLeuIleSer 14
|||||
1495 TTCATAACAACTTTGATATCG 1515

```

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-463-862-5

seq_documentation_block:
; Sequence 5, Application US/08463862
; Patent No. 5776751
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G. et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; NUMBER OF SEQUENCES: 11

```

```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463.862
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,544
; FILING DATE: 16-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6526-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 303..2018
; US-08-463-862-5
;
; alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-674-779-2 x US-08-463-862-5 ..
;
; Align seg 1/1 to: US-08-463-862-5 from: 1 to: 3671
;
; 8 PheileThrLeulleSer 14
; |||||
; 1495 TTCATACAACTTTGATATCG 1515
;
; seq_name: /cgn2.6/ptodata/2/ina/5B_COMB.seq:US-08-461-985-5
;
; seq_documentation_block:
; Sequence 5, Application US/08461985
; Patent No. 5872006
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G.
; APPLICANT: Cobb, Melanie H.
; APPLICANT: Yancopoulos, George D.
; APPLICANT: Nye, Steven
; APPLICANT: Panayotatos, Nikos
; TITLE OF INVENTION: A Family of Map2 Protein Kinases
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461.985
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 800
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,620
; FILING DATE: 03-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6526-123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 303..2018
; US-08-461-985-5
;
; alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-674-779-2 x US-08-461-985-5 ..
;
; Align seg 1/1 to: US-08-461-985-5 from: 1 to: 3671
;
; 8 PheileThrLeulleSer 14
; |||||
; 1495 TTCATACAACTTTGATATCG 1515
;
; seq_name: /cgn2.6/ptodata/2/ina/5B_COMB.seq:US-08-458-887-5
;
; seq_documentation_block:
; Sequence 5, Application US/08458887
; Patent No. 5914261
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G. et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458.887
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 536

```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,544
; FILING DATE: 16-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6526-049
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 303..2018
; US-08-458-887-5

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-458-887-5 ..
Align seg 1/1 to: US-08-458-887-5 from: 1 to: 3671

      8 PheilleThrThrLeuIleSer 14
      |||||||
      1495 TTCATAACAACCTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5

seq_documentation_block:
; Sequence 5, Application US/08932787B
; Patent No. 6277963
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
; FILE REFERENCE: REG 430-A-1
; CURRENT APPLICATION NUMBER: US/08/932,787B
; CURRENT FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 08/469,547
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 CDNA
; US-08-932-787B-5

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-932-787B-5 ..
Align seg 1/1 to: US-08-932-787B-5 from: 1 to: 3671

      8 PheilleThrThrLeuIleSer 14
      |||||||
      1495 TTCATAACAACCTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5

seq_documentation_block:
; Sequence 5, Application US/08932787B
; Patent No. 6277963
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
; FILE REFERENCE: REG 430-A-1
; CURRENT APPLICATION NUMBER: US/08/932,787B
; CURRENT FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 08/469,547
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 CDNA
; US-08-932-787B-5

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-932-787B-5 ..
Align seg 1/1 to: US-08-932-787B-5 from: 1 to: 3671

      8 PheilleThrThrLeuIleSer 14
      |||||||
      1495 TTCATAACAACCTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5

seq_documentation_block:
; Sequence 5, Application US/08932787B
; Patent No. 6277963
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: ANTIBODIES DIRECTED TOWARD EXTRACELLULAR SIGNAL-RELATED
; FILE REFERENCE: REG 430-A-1
; CURRENT APPLICATION NUMBER: US/08/932,787B
; CURRENT FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 08/469,547
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 CDNA
; US-08-932-787B-5

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-932-787B-5 ..
Align seg 1/1 to: US-08-932-787B-5 from: 1 to: 3671

      8 PheilleThrThrLeuIleSer 14
      |||||||
      1495 TTCATAACAACCTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-932-787B-5

seq_documentation_block:
; Sequence 5, Application US/08932787B
; Patent No. 6277963
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; FILE REFERENCE: REG 430-Y-1
; CURRENT APPLICATION NUMBER: US/08/932,012C
; CURRENT FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 08/462,874
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
; PRIOR APPLICATION NUMBER: 07/701,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 CDNA
; US-08-932-012C-5

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-932-012C-5 ..
Align seg 1/1 to: US-08-932-012C-5 from: 1 to: 3671

      8 PheilleThrThrLeuIleSer 14
      |||||||
      1495 TTCATAACAACCTTTGATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-888-818C-5

seq_documentation_block:
; Sequence 5, Application US/08888818C
; Patent No. 6303358
; GENERAL INFORMATION:
; APPLICANT: Boulton et al.
; TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
; FILE REFERENCE: REG 430-V-1
; CURRENT APPLICATION NUMBER: US/08/888,818C
; CURRENT FILING DATE: 1997-07-07
; PRIOR APPLICATION NUMBER: 08/478,985
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/178,488
; PRIOR FILING DATE: 1994-01-07
```

```

; PRIOR APPLICATION NUMBER: 07/01,544
; PRIOR FILING DATE: 1991-05-16
; PRIOR APPLICATION NUMBER: 07/532,004
; PRIOR FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 3671
; TYPE: DNA
; ORGANISM: RAT
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (303)...(2018)
; OTHER INFORMATION: ERK3 CDNA
US-08-888-818C-5

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-888-818C-5 ..
Align seg 1/1 to: US-08-888-818C-5 from: 1 to: 3671

      8 Phe1leThrThrLeu1leSer 14
      |||||
1495 TTCATACAACTTGCATATCG 1515

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5459251-1
seq_documentation_block:
; Patent No. 5459251
; APPLICANT: Tsujimoto, Yoshida;Croce, Carlo A.
; TITLE OF INVENTION: DNA MOLECULES HAVING HUMAN BCL-2 GENE
; SEQUENCES
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,704
; FILING DATE: 18-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 994,941
; FILING DATE: 23-DEC-1992
; APPLICATION NUMBER: 663,010
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: 883,687
; FILING DATE: 09-JUL-1986
; SEQ ID NO:1:
; LENGTH: 4825
5459251-1

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x 5459251-1 ..
Align seg 1/1 to: 5459251-1 from: 1 to: 4825

      31 SerPro1leLysThrProSer 37
      |||||
2446 AGTCCCATCAAACTCCGCT 2466

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-030-096-7
seq_documentation_block:
; Sequence 7, Application US/08030096
; Patent No. 5426041

```

```

; GENERAL INFORMATION:
; APPLICANT: Fabljanski, Steven F.
; APPLICANT: Arnison, Paul G.
; TITLE OF INVENTION: BINARY CRYPTOCYTOTOXIC METHOD OF HYBRID
; TITLE OF INVENTION: SEED PRODUCTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/030,096
; FILING DATE: 22-MAR-1993
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/556,917
; FILING DATE: 20-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA91/00255
; FILING DATE: 22-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/164/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2158..3225, 3663..4046)
; US-08-030-096-7

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x US-08-030-096-7/rev ..
Align seg 1/1 to reverse of: US-08-030-096-7 from: 1 to: 4977

      99 SerGlnLeuLeuThrAla 105
      |||||
374 TCGCAGCTGTTGACCACTGCA 354

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-485A-19
seq_documentation_block:
; Sequence 19, Application US/08465485A
; Patent No. 5831066
; GENERAL INFORMATION:
; APPLICANT: Reed, John
; TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUBSTADT,

```

```

; ADDRESSER: P.C.
; STREET: 1755 S. Jefferson Davis Hwy., Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,485A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,256
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/840,716
; FILING DATE: 21-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/288,692
; FILING DATE: 22-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Fortney, Andrew D.
; REGISTRATION NUMBER: 34,600
; REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (408) 436-2070
; TELEFAX: (408) 436-2075
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5086 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-465-485A-19
;
; alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-674-779-2 x US-08-465-485A-19 ..
;
; Align seg 1/1 to: US-08-465-485A-19 from: 1 to: 5086
;
; 31 SerProLleLysThrProSer 37
; |||||
; 2446 AGTCCCATCAAAACCTCGTCT 2466
;
; seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-365-486A-14
;
; seq_documentation_block:
; Sequence 14, Application US/08365486A
; Patent No. 5834306
; GENERAL INFORMATION:
; APPLICANT: Webster, Keith A.
; APPLICANT: Bishopric, Nanette H.
; TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
; TITLE OF INVENTION: Therapeutic Constructs
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
;
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,486A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 8255-0018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5086 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: human bcl-2 cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1459..2178
;
; US-08-365-486A-14
;
; alignment_scores:
; Quality: 7.00 Length: 7
; Ratio: 1.000 Gaps: 0
; Percent Similarity: 100.000 Percent Identity: 100.000
;
; alignment_block:
; US-09-674-779-2 x US-08-365-486A-14 ..
;
; Align seg 1/1 to: US-08-365-486A-14 from: 1 to: 5086
;
; 31 SerProLleLysThrProSer 37
; |||||
; 2446 AGTCCCATCAAAACCTCGTCT 2466
;
; seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-080-285-19
;
; seq_documentation_block:
; Sequence 19, Application US/09080285
; Patent No. 6040181
; GENERAL INFORMATION:
; APPLICANT: Reed, John
; TITLE OF INVENTION: Regulation of bcl-2 Gene Expression
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Hwy., Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,285
; FILING DATE:
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:

```

```
;
; APPLICATION NUMBER: US 08/465,485
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,256
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/840,716
; FILING DATE: 21-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/288,692
; FILING DATE: 22-DEC-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Fortney, Andrew D.
; REGISTRATION NUMBER: 34,600
; REFERENCE/DOCKET NUMBER: 3335-070-55 CONT
; TELEPHONE: (408) 436-2070
; TELEFAX: (408) 436-2075
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5086 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-080-285-19
;
alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
  US-09-674-779-2 x US-09-080-285-19 ..
  Align seg 1/1 to: US-09-080-285-19 from: 1 to: 5086
  31 SerProIlleLysThrProSer 37
  |||||
  2446 AGTCCCATCAAACTCCGTCT 2466
  seq_name: /cgn2_6/ptodata/2/lna/6B_COMB.seq:US-08-880-342-14
  seq_documentation_block:
  ; Sequence 14, Application US/08880342
  ; Patent No. 6218179
  ; GENERAL INFORMATION:
  ; APPLICANT: Webster, Keith A.
  ; APPLICANT: Bishopric, Nanette H.
  ; APPLICANT: Murphy, Brian
  ; APPLICANT: Laderoute, Keith R.
  ; APPLICANT: Green, Christopher J.
  ; TITLE OF INVENTION: Tissue Specific Hypoxia Regulated
  ; TITLE OF INVENTION: Therapeutic Constructs
  ; NUMBER OF SEQUENCES: 37
  ; CORRESPONDENCE ADDRESS:
  ; ADDRESSEE: Dehlinger & Associates
  ; STREET: 350 Cambridge Avenue, Suite 250
  ; CITY: Palo Alto
  ; STATE: CA
  ; COUNTRY: USA
  ; ZIP: 94306
  ; COMPUTER READABLE FORM:
  ; MEDIUM TYPE: Floppy disk
  ; OPERATING SYSTEM: PC-DOS/MS-DOS
  ; SOFTWARE: Patent In Release #1.0, Version #1.25
  ; CURRENT APPLICATION DATA:
  ; APPLICATION NUMBER: US/08/880,342
  ; FILING DATE: 23-JUN-1997
  ; CLASSIFICATION: 514
  ; PRIOR APPLICATION DATA:
  ; APPLICATION NUMBER: PCT/IB95/00996
;
; FILING DATE: 13-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/365,486
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 8255-0018.30
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5086 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: human bcl-2 cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1459..2178
;
; US-08-880-342-14
;
alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
  US-09-674-779-2 x US-08-880-342-14 ..
  Align seg 1/1 to: US-08-880-342-14 from: 1 to: 5086
  31 SerProIlleLysThrProSer 37
  |||||
  2446 AGTCCCATCAAACTCCGTCT 2466
  seq_name: /cgn2_6/ptodata/2/lna/PCTUS_COMB.seq:PCT-US93-05651-4
  seq_documentation_block:
  ; Sequence 4, Application PC/TUS9305651
  ; GENERAL INFORMATION:
  ; TITLE OF INVENTION: A Gene Which Prevents Programmed Cell Death
  ; NUMBER OF SEQUENCES: 5
  ; COMPUTER READABLE FORM:
  ; MEDIUM TYPE: diskette
  ; CURRENT APPLICATION DATA:
  ; APPLICATION NUMBER: PCT/US93/05651
  ; INFORMATION FOR SEQ ID NO: 4:
  ; SEQUENCE CHARACTERISTICS:
  ; LENGTH: 5086 base pairs
  ; TYPE: nucleic acid
  ; STRANDEDNESS: double
  ; TOPOLOGY: linear
  ; MOLECULE TYPE: cDNA to mRNA
  ; FEATURE:
  ; NAME/KEY: CDS
  ; LOCATION: 1459..2178
  ; OTHER INFORMATION: /product= "Bcl-2"
  ;
  ; PCT-US93-05651-4
;
alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
```


US-09-674-779-2 x PCT-US93-05651-4 ..

Align seg 1/1 to: PCT-US93-05651-4 from: 1 to: 5086

31 SerProlelyThrProSer 37

|||||
2446 AGTCCCATCAAACTCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/PCITUS_COMB.seq:PCT-US93-06251-2

seq_documentation_block:

; Sequence 2, Application PC/TUS9306251

; GENERAL INFORMATION:

; APPLICANT: Wickstrom, Eric and Rife, Jason P.

; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing

; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates

; NUMBER OF SEQUENCES: 93

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

; STREET: 400 Garden City Plaza

; CITY: Garden City

; STATE: NY

; COUNTRY: USA

; ZIP: 11530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/06251

; FILING DATE: 19930630

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Digiglio, Frank S.

; REGISTRATION NUMBER: 31,346

; REFERENCE/DOCKET NUMBER: 8586

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 516-742-4343

; TELEFAX: 516-742-4366

; TELEX: 230 901 SANS UR

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5086 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

PCT-US93-06251-2

alignment_scores:

Quality: 7.00

Ratio: 1.000

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x PCT-US93-06251-2 ..

Align seg 1/1 to: PCT-US93-06251-2 from: 1 to: 5086

31 SerProlelyThrProSer 37

|||||
2446 AGTCCCATCAAACTCGTCT 2466

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-234-186-7

seq_documentation_block:

; Sequence 7, Application US/09234186

; Patent No. 6312947

; GENERAL INFORMATION:

; APPLICANT: Horvitz, H. Robert

; APPLICANT: Hengartner, Michael

; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF A
; TITLE OF INVENTION: GENE WHICH PROTECTS CELLS FROM PROGRAMMED CELL DEATH AND
; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: 01997/201005

; CURRENT APPLICATION NUMBER: US/09/234,186

; CURRENT FILING DATE: 1999-01-20

; EARLIER APPLICATION NUMBER: 07/898,933

; EARLIER FILING DATE: 1992-06-12

; EARLIER APPLICATION NUMBER: 07/927,681

; EARLIER FILING DATE: 1992-08-10

; EARLIER APPLICATION NUMBER: 08/288,295

; EARLIER FILING DATE: 1994-08-10

; EARLIER APPLICATION NUMBER: 08/801,248

; EARLIER FILING DATE: 1997-02-19

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 5086

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1459)...(2178)

US-09-234-186-7

alignment_scores:

Quality: 7.00

Ratio: 1.000

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x US-09-234-186-7 ..

Align seg 1/1 to: US-09-234-186-7 from: 1 to: 5094

31 SerProlelyThrProSer 37

|||||

2454 AGTCCCATCAAACTCGTCT 2474

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5506344-1

seq_documentation_block:

; Patent No. 5506344

; APPLICANT: TSUTSUMOTO, YOSHIMOTO, YOSHIMOTO, CARLO A.

; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR BCL-2 GENE PRODUCT

; NUMBER OF SEQUENCES: 5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/435,193

; FILING DATE: 05-MAY-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 228,704

; FILING DATE: 18-APR-1994

; APPLICATION NUMBER: 994,941

; FILING DATE: 23-DEC-1992

; APPLICATION NUMBER: 663,010

; FILING DATE: 19-MAR-1991

; APPLICATION NUMBER: 883,687

; FILING DATE: 09-JUL-1986

; SEQ ID NO:1

; LENGTH: 5104

5506344-1

alignment_scores:

Quality: 7.00

Ratio: 1.000

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x 5506344-1 ..

Align seg 1/1 to: 5506344-1 from: 1 to: 5104

Tue Sep 17 07:27:53 2002

31 SerProIleLeLysThrProSer 37
|||||
2446 AGTCCCATCAAAACTCCGCTCT 2466

OM of: US-09-674-779-2 to: EST:* out_format : pfs

Date: Sep 16, 2002 7:24 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framed-p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US09674779/runat.12092002.124143.26847/app_query.fasta.1.310
-DB=EST -OFMT=fastap -SUFFIX=oli.rst -GAPOP=4.500 -GAPEXT=0.050
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000
-XGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELOP=6.000
-DELET=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=45
-MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09674779 -CGN1_1_2604 -NCPU=6
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-674-779-2

Query length: 250

Database: EST:*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1623.650000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est2:BF353182	+	9.00	142.58	107.45	579
gb_est2:BF353196	+	9.00	142.58	107.45	579
gb_est2:BF3534072	+	9.00	142.58	112.00	602
gb_gss:AZ374621	+	9.00	141.18	128.69	686
gb_gss:CN5020HD	+	9.00	140.22	145.51	770
gb_gss:CN505M5R	+	9.00	138.25	187.45	977
gb_gss:CN505P19	+	9.00	137.78	192.11	1034
gb_gss:CN505P05	+	9.00	137.25	212.86	1101
gb_gss:CN502KVL	+	9.00	137.21	214.09	1107
gb_gss:BN233428	+	8.00	136.48	235.13	149
gb_est2:BE708381	-	8.00	134.16	316.45	197
gb_est1:AV348302	-	8.00	132.12	411.19	252
gb_est1:AA557886	+	8.00	132.05	414.66	254
gb_est2:BG575333	+	8.00	131.25	459.95	280
gb_est2:BF510695	+	8.00	130.56	502.00	304
gb_est1:BB042909	+	8.00	130.54	503.76	305
gb_est2:BG957563	+	8.00	130.46	509.03	308
gb_est2:BF5109041	+	8.00	130.35	516.06	312
gb_est1:BF1613633	+	8.00	129.03	611.57	366
gb_gss:TA262801Q	+	8.00	128.67	640.04	382
gb_est2:BF934922	-	8.00	128.44	659.66	393
gb_gss:AQ131150	-	8.00	128.23	677.53	403
gb_est1:AV544347	+	8.00	128.08	690.06	410
gb_gss:AQ219969	+	8.00	127.90	706.18	419
gb_est2:BF913055	+	8.00	127.77	718.73	426
gb_est2:BF40672	+	8.00	127.69	725.91	430
gb_est2:BF562471	+	8.00	127.42	751.08	444
gb_gss:AQ600079	+	8.00	127.39	754.68	446
gb_est1:AA215765	+	8.00	127.22	770.89	455
gb_gss:AA509926	+	8.00	127.20	772.69	456
gb_est1:AA874458	+	8.00	127.17	776.29	458
gb_est2:BF565248	+	8.00	127.15	778.10	459
gb_est2:BF513342	+	8.00	127.06	787.12	464
gb_est2:BF513340	+	8.00	127.04	788.92	465
gb_gss:BF105280	-	8.00	126.83	810.59	477
gb_est2:BF1512001	+	8.00	126.62	832.30	489

gb_est2:BF515132 + 8.00 126.18 881.27 516 ! BF515132 BB160017B10C08 Bee
gb_gss:B48301 - 8.00 125.99 903.08 528 ! B48301 RPT111-6P4-TV RPT1-11
gb_est2:BF643242 - 8.00 125.91 912.18 533 ! BF643242 RS4-C01 Sugar beet
gb_est2:BF517071 + 8.00 125.88 915.82 535 ! BF517071 BB160024K20E01 Bee
gb_est1:AI995165 - 8.00 125.74 932.22 544 ! AI995165 701502426 A. thaila
gb_gss:AQ266185 + 8.00 125.22 936.14 579 ! AQ266185 CITBI-E1-2502L22.TF
gb_est2:BF515482 + 8.00 125.21 997.97 580 ! BF515482 BB160019B10B04 Bee

seq_name: gb_est2:BF53182

seq_documentation_block:

LOCUS BF53182 579 bp mRNA linear EST 22-NOV-2000
DEFINITION PM1-HT0628-310800-009-e08 HT0628 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF53182
VERSION BF53182.1 GI:11312256
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 579)

REFERENCE

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL

MEDLINE

COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-HT0628-
310800-009-e08&t3=2000-08-31&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 20

High quality sequence stop: 46.

FEATURES

source

1..579

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0628"

/dev_stage="Adult"

/note="Organ: head_neck; Vector: puc18; Site:1: SmaI;

Site:2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 190 a 131 c 137 g 121 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x BF53182 ..

Align seg 1/1 to: BF53182 from: 1 to: 579

```

85 AlatyrlLeuGlnSerArgLeuGlyAsn 93
|||||
123 GCTTACCTACAAGCGCGTGGGCAAT 149

seq_name: gb_est2:BF353196

seq_documentation_block:
LOCUS BF353196 579 bp mRNA linear EST 22-NOV-2000
DEFINITION PM1-HT0629-290800-009-e08 HT0629 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF353196
VERSION BF353196.1 GI:11312270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 579)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Simpson,D.H.,
Goldman,G.H., Carvalheiro,A.F., Matsukuma,A., Balg,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&t2=PM1-HT0629-
290800-009-e08&t3=2000-08-29&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 46.
FEATURES
Location/Qualifiers
source 1..579
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0629"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 190 a 131 c 137 g 121 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BF353196 ..
Align seg 1/1 to: BF353196 from: 1 to: 579

85 AlatyrlLeuGlnSerArgLeuGlyAsn 93
|||||
123 GCTTACCTACAAGCGCGTGGGCAAT 149

seq_name: gb_est2:BF344072

seq_documentation_block:
LOCUS BE344072 602 bp mRNA linear EST 17-JUL-2000
DEFINITION EST409234 potato stolon, Cornell University Solanum tuberosum cDNA
clone CSTA28A10, mRNA sequence.
ACCESSION BE344072
VERSION BE344072.1 GI:9253604
KEYWORDS EST.
SOURCE Solanum tuberosum
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 602)
AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan
,B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
,B.
TITLE Generation of ESTs from potato swelling stolons
JOURNAL Unpublished (1999)
COMMENT Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195
Email: cdna@resgen.com
5 prime sequence.
FEATURES
Location/Qualifiers
source 1..602
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone_lib="CSTA28A10"
/clone_lib="potato stolon, Cornell University"
/tissue_type="axillary buds of stem explants, swelling
stolons"
/dev_stage="1 to 3 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; RNA was supplied by Christian Bachem & Beatrix
Horvath(Laboratory of Plant Breeding, Dept. of Plant
Sciences, Wageningen University, The Netherlands). Total
RNA was isolated from developing axillary buds of potato
nodal stem cuttings cultured on medium for the
introduction of tuber formation as described in Bachem et
al. (Plant Journal 1996). Tissue samples were taken of
stages corresponding to growing stolons and the early
stages of tuber formation."
BASE COUNT 183 a 117 c 127 g 175 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BE344072 ..
Align seg 1/1 to: BE344072 from: 1 to: 602

99 SerGlnLeuThrAlaArgSer 107
|||||
497 TCCCACTGCTACACAGCGGTCT 523

seq_name: gb_gss:AZ374621

seq_documentation_block:
LOCUS AZ374621 686 bp DNA linear GSS 02-OCT-2000
DEFINITION 1M0127G01R Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0127G01 R, DNA sequence.
ACCESSION AZ374621
VERSION AZ374621.1 GI:10488321
KEYWORDS GSS.

```

```

SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 686)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0127 row: G column: 01
            Seq primer: CACACAGGAACAGCATGACC
            Class: plasmid ends
            High quality sequence stop: 686.
FEATURES   Location/Qualifiers
            source
            1..686
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UGGCM10127G01"
            /clone_lib="Mouse 10kb plasmid UGGCM1 library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
            /note="Vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of pWB42 (gil14732114|gb|AF129072.1), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."
BASE COUNT   174 a 168 c 110 g 234 t
ORIGIN

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AZ374621
..
Align seg 1/1 to: AZ374621 from: 1 to: 686
7 TyrPheIleThrThrLeuIleSerSer 15
178 TACTTCATACTACATTAATCATGTTCA 204
seq_name: gb_gss:CNS020HD

```

```

seq_documentation_block:
LOCUS      CNS020HD
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
            22024 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL175594.1 GI:7813651
VERSION    GSS: genome survey sequence.
KEYWORDS   Tetraodon nigroviridis.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 770)
AUTHORS     Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
TITLE       Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
AUTHORS     2 (bases 1 to 770)
            Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
TITLE       Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
JOURNAL     Unpublished
AUTHORS     3 (bases 1 to 770)
            Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES   Location/Qualifiers
            source
            1..770
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="222024"
            /clone_lib="G"
            /note="Genoscope sequence ID : COAG222BHI2LPI-end : T7"
BASE COUNT   233 a 140 c 179 g 213 t 5 others
ORIGIN

alignment_scores:
    Quality: 9.00      Length: 9
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x CNS020HD/rev
..
Align seg 1/1 to reverse of: CNS020HD from: 1 to: 770
116 TyrGluLeuProGluHisLeuTrp 124
|||||
581 TACCAGCTGCCTCCAGAACACACTTGTGG 555
seq_name: gb_gss:CNS05M5R
seq_documentation_block:
LOCUS      CNS05M5R
DEFINITION Tetraodon nigroviridis genome survey sequence T3 end of clone
            02908 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL343656
VERSION    AL343656.1 GI:8237426
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis.

```

REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
TITLE	1 (bases 1 to 977)
JOURNAL AUTHORS	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL AUTHORS	2 (bases 1 to 977)
TITLE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL AUTHORS	Unpublished
TITLE	3 (bases 1 to 977)
TITLE	Genoscope.
TITLE	Direct Submission
JOURNAL AUTHORS	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
TITLE	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
TITLE	http://www.genoscope.cns.fr/tetraodon.
FEATURES	Location/Qualifiers
source	1..977
BASE COUNT	272 a 263 c 198 g 217 t 27 others
ORIGIN	
alignment_scores:	
Quality:	9.00 Length: 9
Ratio:	1.000 Gaps: 0
Percent Similarity:	100.000 Percent Identity: 100.000
alignment_block:	
US-09-674-779-2 x CNS05M5R	..
Align seg 1/1 to: CNS05M5R from: 1 to: 977	
seq_name: gb_gss:CNS05P19	
seq_documentation_block:	
LOCUS	1034 bp DNA linear GSS 26-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence T3 end of clone 035C15 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL347382
VERSION	AL347382.1 GI:8241152
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE AUTHORS	1 (bases 1 to 1034)
TITLE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL AUTHORS	2 (bases 1 to 1034)
TITLE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL AUTHORS	Unpublished
TITLE	3 (bases 1 to 1034)
TITLE	Genoscope.
TITLE	Direct Submission
JOURNAL AUTHORS	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
TITLE	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at
TITLE	http://www.genoscope.cns.fr/tetraodon.
FEATURES	Location/Qualifiers
source	1..1034
BASE COUNT	285 a 284 c 213 g 230 t 22 others
ORIGIN	
alignment_scores:	
Quality:	9.00 Length: 9
Ratio:	1.000 Gaps: 0
Percent Similarity:	100.000 Percent Identity: 100.000
alignment_block:	
US-09-674-779-2 x CNS05P19	..
Align seg 1/1 to: CNS05P19 from: 1 to: 1034	
seq_name: gb_gss:CNS05FU5	
seq_documentation_block:	
LOCUS	1101 bp DNA linear GSS 26-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 005M17 of library A from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION	AL335462
VERSION	AL335462.1 GI:8229220
KEYWORDS	GSS; genome survey sequence.
SOURCE	Tetraodon nigroviridis
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE AUTHORS	1 (bases 1 to 1101)
TITLE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL AUTHORS	2 (bases 1 to 1101)
TITLE	Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL AUTHORS	Unpublished
TITLE	3 (bases 1 to 1101)
TITLE	Genoscope.

TITLE Direct Submission
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source
 1..1101
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="005M17"
 /clone_lib="A"
 /note="Genoscope sequence ID : COAA005AG09C1-end : T7"

BASE COUNT 298 a 311 c 221 g 264 t 7 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x CNS05FU5 ..

Align seg 1/1 to: CNS05FU5 from: 1 to: 1101

116 TyrGlnLeuProGluHisLeuTrp 124

|||||
 729 TACCAGCTGCCTCCAGACACTTGTTGG 755

seq_name: gb_gss:CNS02KVL

seq_documentation_block:

LOCUS CNS02KVL 1107 bp DNA linear GSS 14-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
 146C02 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL202026

VERSION AL202026.1 GI:7860371

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1107)

AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1107)

AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 1107)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source
 1..1107
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"

/clone="146C02"
 /clone_lib="G"
 /note="Genoscope sequence ID : COAG146BB01LP1-end : T7"
 BASE COUNT 300 a 221 c 270 g 310 t 6 others
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x CNS02KVL/rev ..

Align seg 1/1 to reverse of: CNS02KVL from: 1 to: 1107

116 TyrGlnLeuProGluHisLeuTrp 124

|||||
 481 TACCAGCTGCCTCCAGACACTTGTTGG 455

seq_name: gb_gss:BH233428

seq_documentation_block:

LOCUS BH233428 149 bp DNA linear GSS 08-NOV-2001
 DEFINITION 1006173F09.2EL_x1 1006 - RescueMu Grid G Zea mays genomic, DNA
 sequence.

ACCESSION BH233428

VERSION BH233428.1 GI:16839654

KEYWORDS GSS.

SOURCE Zea mays.

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea..

REFERENCE 1 (bases 1 to 149)

AUTHORS Walbot,V.

TITLE Maize genomic sequences found using engineered RescueMu transposon

JOURNAL Unpublished (2001)

COMMENT Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Possible ligation site of ends cut by 2 different endonucleases.

Reverse complemented post-ligation sequence from source sequence.

Plate: 1006173 row: 5

Class: transposon-tagged.

Location/Qualifiers

1..149

/organism="Zea mays"

/cultivar="mixed background W23/A188/B73"

/db_xref="taxon:4577"

/clone_lib="1006 - RescueMu Grid G"

/tissue_type="leaf"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: leaf; Vector: RescueMu (engineered from

pBlueScript backbone); Site.1: BamHI; Site.2: BglII;

RescueMu is a 4.9 kb, modified maize Mu transposon

designed to allow plasmid rescue from total genomic DNA.

Mu elements insert preferentially into transcription

units. For more information on RescueMu, go to the web

site 'www.zmdb.iastate.edu' and follow the links for

'RescueMu.' Grid G was grown at Stanford in 2000. DNA was

extracted from leaf punches, double digested using BamHI

and BglII, and ligated to form circular plasmids. DH10B

cells were transformed and then screened on LB plates with

ampicillin."

BASE COUNT 32 a 36 c 21 g 60 t
 ORIGIN

Tue Sep 17 07:27:54 2002

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x BH233428 ..
Align seq 1/1 to: BH233428 from: 1 to: 149

170 HisLeuThrAsnSerAlaIleasp 177

28 CACCTAACCACTCGGCTATAGAC 51

seq_name: gb_est2:BE708381

seq_documentation_block:

LOCUS BE708381 197 bp mRNA linear EST 12-SEP-2000
DEFINITION MR0-HR0559-060600-030-c10 HT0559 Homo sapiens CDNA, mRNA sequence.

ACCESSION BE708381

VERSION BE708381.1 GI:10096646

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

1 (bases 1 to 197)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-MR0-HT0559-060

600-030-c10&t3=2000-06-06&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 22

High quality sequence stop: 137.

FEATURES

Location/Qualifiers

1..197

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="HT0559"

/dev_stage="Adult"

/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 55 a 46 c 47 g 49 t

ORIGIN

alignment_scores: 8.00 Length: 8

alignment_block:

US-09-674-779-2 x BE708381/rev ..

Align seq 1/1 to reverse of: BE708381 from: 1 to: 197

58 SerHisValGlyLeuGlnAlaHis 65

|||||

105 TCACACCTTGGACTGCACGACAC 82

seq_name: gb_est1:AV348302

seq_documentation_block:

LOCUS AV348302 252 bp mRNA linear EST 12-NOV-1999
DEFINITION AV348302 RIKEN full-length enriched, adult male olfactory bulb Mus

musculus cDNA clone 6430704H12 3', mRNA sequence.

ACCESSION AV348302

VERSION AV348302.1 GI:6389361

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 252)

Konno, H., Alzawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,

Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F.,

Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai

, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,

Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,

Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata

, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,

Suzuki, H., Takahashi, F., Tatenio, M., Tomimaga, N., Tsunoda, Y.,

Washiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,

Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE

RIKEN Mouse ESTs (Konno, H., et al. 1999)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,

Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki

, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)

Itoh, M., Kitsumi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki

, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

Location/Qualifiers

1..252

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_lib="RIKEN full-length enriched, adult male

olfactory bulb"

/sex="male"

/tissue_type="olfactory brain"

/dev_stage="adult"

FEATURES

source


```

/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATTAATCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI"

```

```

BASE COUNT      67 a  48 c  71 g  66 t
ORIGIN

```

```

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:        0
  Percent similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:

```

```

US-09-674-779-2 x AV348302/rev ..

```

```

Align seg 1/1 to reverse of: AV348302 from: 1 to: 252

```

```

219 AlalHeHisLeuAspThrGlnGly 226
134 GCTATACACCTGGATACCCAGGC 111

```

```

seq_name: gb_est1:AA557886

```

```

seq_documentation_block:

```

```

LOCUS      AA557886      254 bp mRNA linear EST 09-SEP-1997
DEFINITION n164f02.s1 NCI_CGAP_Pr4.1 Homo sapiens cDNA clone IMAGE:1045467,
            mRNA sequence.

```

```

ACCESSION  AA557886

```

```

VERSION    AA557886.1 GI:2328363

```

```

KEYWORDS   EST.

```

```

SOURCE     human.

```

```

ORGANISM   Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE  1 (bases 1 to 254)

```

```

AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

```

TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

```

```

JOURNAL   Tumor Gene Index

```

```

COMMENT   Unpublished (1997)

```

```

Contact: Robert Strausberg, Ph.D.

```

```

Email: cgapbs-r@mail.nih.gov

```

```

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.

```

```

, Michael R. Emmert-Buck, M.D., Ph.D.

```

```

cDNA Library Preparation: David B. Krizman, Ph.D.

```

```

DNA Sequencing by: Greg Lennon, Ph.D.

```

```

Clone distribution: NCI-CGAP clone distribution information can be

```

```

found through the I.M.A.G.E. Consortium/LLNL at:

```

```

www-bio.llnl.gov/bbrp/image/image.html

```

```

Insert Length: 359 Std Error: 0.00

```

```

Seq primer: -40m13 fwd. ET from Amersham.

```

```

Location/Qualifiers

```

```

1..254

```

```

/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"

```

```

/clone="IMAGE:1045467"

```

```

/clone_lib="NCI_CGAP_Pr4.1"

```

```

FEATURES

```

```

Source

```

```

FEATURES

```

```

Source

```

```

Location/Qualifiers

```

```

1..250

```

```

/organism="Rattus norvegicus"

```

```

/strain="Sprague-Dawley"

```

```

/sex="male"
/tissue_type="prostatic intraepithelial neoplasia - high
grade"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: pAMP10; mRNA made from
prostatic intraepithelial neoplasia (high-grade), cDNA
made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp.
cDNA Library Preparation: David B. Krizman, Ph.D.
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383. cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing
Center"

```

```

BASE COUNT      76 a  63 c  51 g  64 t
ORIGIN

```

```

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:        0
  Percent similarity: 100.000 Percent Identity: 100.000

```

```

alignment_block:

```

```

US-09-674-779-2 x AA557886/rev ..

```

```

Align seg 1/1 to reverse of: AA557886 from: 1 to: 254

```

```

211 GlyLeuGlyLeuTyraAlaThrGly 218
|||||
38 GGTGGGGGCTTTATGCCACTGGG 15

```

```

seq_name: gb_est2:BG375333

```

```

seq_documentation_block:

```

```

LOCUS      BG375333

```

```

DEFINITION UI-R-CV1-bsz-b-12-0-UI.s1 UI-R-CV1 Rattus norvegicus cDNA clone
            UI-R-CV1-bsz-b-12-0-UI 3', mRNA sequence.

```

```

ACCESSION  BG375333

```

```

VERSION    BG375333.1 GI:13299805

```

```

KEYWORDS   EST.

```

```

SOURCE     Norway rat.

```

```

ORGANISM   Rattus norvegicus

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE  1 (bases 1 to 280)

```

```

AUTHORS   Bonaldo, M.F., Lennon, G. and Soares, M.B.

```

```

TITLE     Normalization and subtraction: two approaches to facilitate gene

```

```

JOURNAL   discovery

```

```

MEDLINE   97044477

```

```

COMMENT   Genome Res. 6 (9), 791-806 (1996)

```

```

Contact: Soares, MB

```

```

Program for Rat Gene Discovery and Mapping

```

```

University of Iowa

```

```

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

```

```

Tel: 319 335 8250

```

```

Fax: 319 335 9565

```

```

Email: msoares@blue.weeg.uiowa.edu

```

```

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized rat eye library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com) The following repetitive elements were
found in this cDNA sequence: 1-35, >AT-rich#Low_complexity
Seq primer: M13 Forward
POLYA=Yes.

```

```

/db_xref="taxon:10116"
/clone="UI-R-CV1-bsz-b-12-0-UI"
/clone_lib="UI-R-CV1"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CV1
library is a normalized library constructed from rat eye
tissue. For a detailed description of the library from
which this clone was derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LJB=UI-R-CV1
TAG_TISSUE=rat eye
TAG_SEQ=CAGCC" 52 g 97 t
BASE COUNT 75 a 56 c
ORIGIN

```

```

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BG375333 ..
Align seg 1/1 to: BG375333 from: 1 to: 280
14 SerSerMetLeuValAlaCysSer 21
|||||
252 AGTCATGTTAGTTCGTCGTCAGC 275

```

```
seq_name: gb_est2:B1510695
```

```

seq_documentation_block:
LOCUS B1510695 304 bp mRNA linear EST 29-AUG-2001
DEFINITION B160003A20C06 Bee Brain Normalized Library, BB16 Apis mellifera
CDNA clone B160003A20C06 5', mRNA sequence.
ACCESSION B1510695
VERSION B1510695.1 GI:15361069
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
; Apoidea; Apidae; Apis.
Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
Smoller,D. and Robinson,G.E.
An Expressed Sequence Tag Resource for Studies of Brain and
Behavior in the Honey Bee
Unpublished (2001)
Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TATACGACTCTACTATAGG
BACKWARD: ATTACCTCTACTTAAG
Insert Length: 304 Std Error: 0.00
Plate: B160003A20 row: C column: 06
Seq primer: AGCGGATCAAAATTCACACAGGA
High quality sequence stop: 304.
Location/Qualifiers

```

```
source
```

```

1. 304
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="B160003A20C06"
/clone_lib="Bee Brain Normalized Library, BB16"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT73-Pac; Site 1: EcoRI;
Site 2: NotI; The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
BASE COUNT 82 a 60 c 72 g 90 t
ORIGIN

```

```

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x B1510695 ..
Align seg 1/1 to: B1510695 from: 1 to: 304
208 GlnAsnPheGlyLeuGlyLeuTyr 215
|||||
75 CAGAAATTCGATTAGGCTTTAT 98

```

```
seq_name: gb_est1:BB042909
```

```

seq_documentation_block:
LOCUS BB042909 305 bp mRNA linear EST 23-JUN-2000
DEFINITION BB042909 RIKEN full-length enriched, 13 days embryo male testis Mus
musculus cDNA clone 6030468A09 3', mRNA sequence.
ACCESSION BB042909
VERSION BB042909.1 GI:8449295
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 305)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
,Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
, Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
, Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
,Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resesc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/

```

```

TITLE
COMMENT

```

```
FEATURES
```

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermotabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Nishiyama,Y., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1..305
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, 13 days embryo male testis"
/sex="male"
/tissue_type="testis"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTTCGAGTTAAATTAATCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT
ORIGIN

63 a 82 c 50 g 110 t
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x BB042909 from: 1 to: 305
Align seg 1/1 to: BB042909 from: 1 to: 305

137 LeuLysSerArgGlyIleLeuPro 144
|||||
78 CTCAAGTCCAGAGGAATCTGCCT 101

seq_name: gb_est2:BG957563

seq_documentation_block:

LOCUS BG957563 308 bp mRNA linear EST 12-JUN-2001
DEFINITION CM2-CT0662-020301-696-d09 CT0662 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG957563
VERSION BG957563.1 GI:14375734
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 308)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

20202663

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=CM2&t2=CM2-CT0662-020301-696-d09&t3=2001-03-02&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 11

High quality sequence stop: 308.

Location/Qualifiers

1..308

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CT0662"

/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORSTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 64 a 105 c 83 g 56 t

ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x BG957563/rev ..
Align seg 1/1 to reverse of: BG957563 from: 1 to: 308

101 LeuLeuThrThrAlaArgSerTrp 108
|||||
250 CTTCTAACAACTGGCGTTCATGS 227

seq_name: gb_est2:BI509041

seq_documentation_block:

LOCUS BI509041 312 bp mRNA linear EST 29-AUG-2001
DEFINITION BB170015B20G01 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170015B20G01 5', mRNA sequence.
ACCESSION BI509041
VERSION BI509041.1 GI:15359415
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
; Apoidea; Apidae; Apis.
1 (bases 1 to 312)

AUTHORS Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L., Smoller,D. and Robinson,G.E.
TITLE An Expressed Sequence Tag Resource for Studies of Brain and Behavior in the Honey Bee
JOURNAL Unpublished (2001)
COMMENT Contact: Gene E. Robinson
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0309
 Fax: 217 244 3499
 Email: generobi@life.uiuc.edu
 This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR PRIMERS
 FORWARD: TAATACGCTCCTACTATAGGG
 BACKWARD: ATTAACCTCCTACTAAG
 Insert Length: 312 Std Error: 0.00
 Plate: BB170015B20 row: G column: 01
 Seq primer: AGCGGATACCAATTTCACACAGGA
 High quality sequence stop: 312.
 Location/Qualifiers
 1..312
 /organism="Apis mellifera"
 /strain="mixed strains of European bees, predominantly A.m. ligustica"
 /db_xref="taxon:7460"
 /clone="BB170015B20G01"
 /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
 /sex="female"
 /tissue_type="brain"
 /dev_stage="adult worker honey bee"
 /lab_host="DH108"
 /note="Organ: brain; Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups."

FEATURES
 source

BASE COUNT 83 a 60 c 73 g 96 t
ORIGIN
 alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000
 alignment_block:
 US-09-674-779-2 x BI509041 ..
 Align seg 1/1 to: BI509041 from: 1 to: 312

208 GlnAsnPheGlyLeuGlyLeuTyr 215
 |||||
 73 CAGATTTCGGATTAGGCTTTAT 96
 seq_name: gb_est1:AI613633
 seq_documentation_block:
 LOCUS AI613633 366 bp mRNA linear EST 21-APR-1999
 DEFINITION vq30c10.v1 Soares_mammary_gland_NDMMG Mus musculus cDNA clone IMAGE:862866 5', mRNA sequence.
 ACCESSION AI613633
 VERSION AI613633.1 GI:4622800
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 366)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapps@mail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:506954

FEATURES
 source

1..366
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:862866"
 /clone_lib="Soares_mammary_gland_NDMMG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH108"
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."

BASE COUNT 92 a 114 c 108 g 52 t
ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AI613633/rev ..
 Align seg 1/1 to reverse of: AI613633 from: 1 to: 366

101 LeuLeuThrAlaArgSerTrp 108
 |||||
 190 CTCTAACAACTGCGGCTCATGG 167

seq_name: gb_gss:TA262B01Q

seq_documentation_block:
 LOCUS TA262B01Q 382 bp DNA linear GSS 13-DEC-2000
 DEFINITION T. brucei sheared genomic DNA clone 262b01, reverse sequence, genomic survey sequence.

ACCESSION AL487669
 VERSION AL487669.1 GI:11850705
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 382)
 AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 366)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapps@mail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:506954

This read is a RESEQUENCE of a previously sequenced mouse clone
 this read has been verified (found to hit its original self in the
 correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 365.
 Location/Qualifiers
 1..366
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:862866"
 /clone_lib="Soares_mammary_gland_NDMMG"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DH108"
 /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
 constructed and normalized by Bento Soares and M.Fatima
 Bonaldo."

BASE COUNT 92 a 114 c 108 g 52 t
ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AI613633/rev ..
 Align seg 1/1 to reverse of: AI613633 from: 1 to: 366

101 LeuLeuThrAlaArgSerTrp 108
 |||||
 190 CTCTAACAACTGCGGCTCATGG 167

seq_name: gb_gss:TA262B01Q

seq_documentation_block:
 LOCUS TA262B01Q 382 bp DNA linear GSS 13-DEC-2000
 DEFINITION T. brucei sheared genomic DNA clone 262b01, reverse sequence, genomic survey sequence.

ACCESSION AL487669
 VERSION AL487669.1 GI:11850705
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei

REFERENCE 1 (bases 1 to 382)
 AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,

TITLE
JOURNAL
COMMENT

Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, the Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).

Email: nelsaudet@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/projects/T-brucei/>.

FEATURES
 source
 1. 382
 /organism="Trypanosoma brucei"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="262B01"

BASE COUNT
 99 a 79 c 61 g 142 t 1 others

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x TA262B01Q ..
 Align seg 1/1 to: TA262B01Q from: 1 to: 382

123 LeuTrpGlyGlnIleValProThr 130
 |||||
 53 TTGTGGGTCAAAATGCTCTACC 76

seq_name: gb_est2:BF934922
seq_documentation_block:
LOCUS BF934922 393 bp mRNA linear EST 22-JAN-2001
DEFINITION IL2-NT0201-231200-323-H06 NT0201 Homo sapiens CDNA, mRNA sequence.
ACCESSION BF934922
VERSION BF934922.1 GI:12352246
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

TITLE
JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL2&t2=IL2-NT0201-231200-323-H06&t3=2000-12-23&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence stop: 392.

FEATURES
 source

Location/Qualifiers
 1. 393
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0201"
 /dev_stage="Adult"

/note="Organ: nervous_tumor; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 121 a 69 c 82 g 120 t 1 others
ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BF934922/rev ..
 Align seg 1/1 to reverse of: BF934922 from: 1 to: 393

101 LeuLeuThrAlaArgSerTrp 108
 |||||
 331 CTTTGGACCACTGCTCGTCTCG 308

seq_name: gb_gss:AQ131150

seq_documentation_block:
LOCUS AQ131150 403 bp DNA linear GSS 23-SEP-1998
DEFINITION HS_3036_AL_F06_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3036 Col-11 Row-K, DNA sequence.

ACCESSION AQ131150
VERSION AQ131150.1 GI:3508316
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 403)

REFERENCE
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589

COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3036 row: K column: 11
 Class: BAC ends
 High quality sequence stop: 403.

FEATURES
 source

Location/Qualifiers
 1. 403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

Tue Sep 17 07:27:54 2002

```

/clone="Plate-3036 Col-11 Row-K"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelobAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT      95 a 112 c 53 g 143 t
ORIGIN

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x A0131150/rev ..
  Align seg 1/1 to reverse of: A0131150 from: 1 to: 403

137 LeuLysSerArgGlyIleuPro 144
|||||
322 CTCAAAGCAGGGGATTTGCCA 299

seq_name: gb_est1:AV544347

seq_documentation_block:
LOCUS      AV544347      410 bp      mRNA      linear      EST 07-SEP-2000
DEFINITION AV544347 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            CDNA clone R240g10F 3', mRNA sequence.
ACCESSION  AV544347
VERSION     AV544347.1 GI:8715761
KEYWORDS   EST.
SOURCE      Chae cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 410)
AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
JOURNAL   DNA Res. 7, 175-180 (2000)
MEDLINE   20363093
COMMENT   Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
            Location/Qualifiers
            1..410
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /clone="R240g10F"
                /clone_lib="Arabidopsis thaliana roots Columbia"
                /tissue_type="roots"
                /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                XhoI"
BASE COUNT      128 a 106 c 56 g 120 t
ORIGIN

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AV544347 ..
  Align seg 1/1 to: AV544347 from: 1 to: 410

```

```

144 ProAlaAsnThrGlnIleArgSer 151
|||||
152 CCGGCCAATACAGATCCGATCC 175

```

seq_name: gb_gss:A0219969

```

seq_documentation_block:
LOCUS      A0219969      419 bp      DNA      linear      GSS 19-SEP-1998
DEFINITION HS_3251_B2_G07_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate-3251 Col-14 Row=N, DNA sequence.
ACCESSION  A0219969
VERSION     A0219969.1 GI:3633582
KEYWORDS   GSS.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 419)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.

```

```

TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3251 row: N column: 14
            Class: BAC ends
            High quality sequence stop: 419.

```

```

FEATURES
            Location/Qualifiers
            1..419
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate-3251 Col-14 Row=N"
                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
                E-Coli DH10B"
BASE COUNT      117 a 89 c 80 g 126 t 7 others
ORIGIN

```

```

alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x A0219969 ..
  Align seg 1/1 to: A0219969 from: 1 to: 419

```

```

230 TrpGlyAlaGlnPheSerGluThr 237
|||||
206 TGGGGAGCTCAGTTCTCAGAAACT 229

```

seq_name: gb_est2:BG913055

```

seq_documentation_block:
LOCUS      BG913055      426 bp      mRNA      linear      EST 05-JUN-2001
DEFINITION 602806433F1 NCI CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4938513
            5', mRNA sequence.
ACCESSION  BG913055
VERSION     BG913055.1 GI:14293531
KEYWORDS   EST.
SOURCE      human.

```

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 426)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LLAM10874 row: a column: 10
 High quality sequence stop: 426.

FEATURES
 source
 1..426
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4938513"
 /clone_lib="NCI_CGAP_Brn67"
 /tissue_type="anaplastic oligodendroglioma with 1p/19q
 loss"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.3 Kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 97 a 106 c 82 g 141 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x BG913055/rev ..

Align seg 1/1 to reverse of: BG913055 from: 1 to: 426

167 MetSerLysHisLeuThrAsnSer 174
 |||||
 253 ATGAGCAACATCTGACAAATTCA 230

seq_name: gb_est2:R40672

seq_documentation_block:
 LOCUS R40672 430 bp mRNA linear EST 22-MAY-1995
 DEFINITION yf79e11.s1 Soares infant brain lN1B Homo sapiens cDNA clone
 IMAGE:28343 3', mRNA sequence.
 ACCESSION R40672
 VERSION R40672.1 GI:821010
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 430)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 COMMENT On May 5, 1995 this sequence version replaced gi:798288.
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TITLE
 JOURNAL
 COMMENT

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1430
 High quality sequence stops: 251 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1430 Std Error: 0.00
 Seq primer: Promega -21m13
 High quality sequence stop: 251.

FEATURES
 source
 1..430
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="GDB:400690"
 /db_xref="taxon:9606"
 /clone="IMAGE:28343"
 /clone_lib="Soares infant brain lN1B"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: Lfamid BA; Site_1: Not
 I; Site_2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5,
 AACTGGAACAAATTCGCGCCGAGGAATTTTTTTTTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the Lfamid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Patima Bonaldo."

BASE COUNT 115 a 81 c 84 g 146 t 4 others
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x R40672/rev ..

Align seg 1/1 to reverse of: R40672 from: 1 to: 430

24 IleProThrAsnProGlnValSer 31
 |||||
 374 ATACCCACNAATCCACAGTGAGT 351

seq_name: gb_est2:BI513696

seq_documentation_block:
 LOCUS BI513696 444 bp mRNA linear EST 29-AUG-2001
 DEFINITION BB160013BI10D03 Bee Brain Normalized Library, BB16 Apis mellifera
 cDNA clone BB160013BI10D03 5', mRNA sequence.
 ACCESSION BI513696
 VERSION BI513696.1 GI:15364070
 KEYWORDS EST.
 SOURCE honeybee.

ORGANISM Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
 ; Apoidea; Apidae; Apis.
 REFERENCE 1 (bases 1 to 444)
 AUTHORS Whitfield, C.W., Soares, B., Robertson, H.M., Pardinas, J., Liu, L.,
 Smoller, D. and Robinson, G.E.
 TITLE An Expressed Sequence Tag Resource for Studies of Brain and
 Behavior in the Honey Bee
 JOURNAL Contact: Gene E. Robinson
 COMMENT Unpublished (2001)
 Department of Entomology
 University of Illinois
 505 S. Goodwin Ave., Urbana, IL 61801, USA
 Tel: 217 265 0309
 Fax: 217 244 3499

Email: generobi@life.uiuc.edu
 This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
 PCR Primers
 FORWARD: TAATACGACTCACTATAGG
 BACKWARD: ATTAACCCCTCACTAAAG
 Insert Length: 444 Std Error: 0.00
 Plate: BB160013B10 row: D column: 03
 Seq primer: AGCGATACAAATTCACACAGGA
 High quality sequence stop: 444.

FEATURES

source

```
1. .444
Location/Qualifiers
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB160013B10P03"
/clone_lib="Bee Brain Normalized Library, BB16"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3-Pac; Site:1: EcoRI;
Site:2: NotI; The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
```

BASE COUNT 149 a 69 c 95 g 131 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x B1513696 ..

Align seg 1/1 to: B1513696 from: 1 to: 444

208 GlnAsnPhcglyLeuGlyLeuTyr 215

63 CAAAATTTGGTTGGCATTATAT 86

seq_name: gb_est2:BF562471

seq_documentation_block:
 LOCUS BF562471 446 bp mRNA linear EST 12-DEC-2000
 DEFINITION UI-R-BSO-ans-e-07-0-UI.r1 UI-R-BSO Rattus norvegicus cDNA clone
 UI-R-BSO-ans-e-07-0-UI 5', mRNA sequence.

ACCESSION BF562471
 VERSION BF562471.1 GI:11672249
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 446)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL 97044477

MEDLINE

COMMENT

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250
 Fax: 319 335 9565

Email: mscoates@blue.weeg.uiowa.edu
 CDNA Library Preparation: M.B. Soares Lab Clone distribution:
 clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at
 LLNL (info@image.llnl.gov). IMAGE ID= 1801788
 Seq primer: M13 Forward.

FEATURES

source

```
Location/Qualifiers
1. .446
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BSO-ans-e-07-0-UI"
/clone_lib="UI-R-BSO"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker; Site:1: Not I; Site:2: Eco RI; The UI-R-BSO
library is derived from 13 dpc whole embryo tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
ratest.eng.uiowa.edu."
```

BASE COUNT 101 a 118 c 118 g 109 t
 ORIGIN

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x BF562471/rev ..

Align seg 1/1 to reverse of: BF562471 from: 1 to: 446

124 TrpGlyGlnIleValProThrLeu 131

429 TGGGGCAAAATGTGCCAACACTG 406

seq_name: gb_gss:AQ600079

seq_documentation_block:

LOCUS AQ600079 446 bp DNA linear GSS 10-JUN-1999
 DEFINITION HS.5354.B2.G05.SP6E.RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate-930 Col=10 Row=N, DNA sequence.

ACCESSION AQ600079
 VERSION AQ600079.1 GI:5059996
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 446)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

REFERENCE 1 (bases 1 to 446)
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pletier de Jong
 (pletier@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 930 row: N column: 10
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 446.
 Location/Qualifiers
 1. 446

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT      137 a   89 c   101 g   119 t
ORIGIN

```

alignment_scores:

```

Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-674-779-2 x AQ600079 ..

Align seg 1/1 to: AQ600079 from: 1 to: 446

239 SerilleCysArgHisValLeuPro 246

|||||

183 TCTATTTCGGCGACGTCTACCC 206

seq_name: gb_est1:AW215765

seq_documentation_block:

```

LOCUS      AW215765      455 bp      mRNA      linear      EST 06-DEC-1999
DEFINITION up09g09.y1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:2651584 5',
            mRNA sequence.
ACCESSION  AW215765
VERSION    AW215765.1 GI:6526460
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone Distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

```

MGI:1032036

Seq primer: -40RP from Gibco

High quality sequence stop: 450.

FEATURES

source

```

Location/Qualifiers
1. 455
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:2651584"

```

```

/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; transgenic model WNT-1, expression driven by
MTMV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT      77 a   144 c   138 g   96 t
ORIGIN

```

alignment_scores:

```

Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-674-779-2 x AW215765/rev ..

Align seg 1/1 to reverse of: AW215765 from: 1 to: 455

139 SerArgGlyLeuProAlaAsn 146

|||||

383 TCACGAGGGATCTGCCAGCCAAC 360

seq_name: gb_est1:AA509926

seq_documentation_block:

```

LOCUS      AA509926      456 bp      mRNA      linear      EST 08-JUL-1997
DEFINITION vg30c10.r1 Soares_mammary_gland_NBMG Mus musculus cDNA clone
            IMAGE:862866 5', mRNA sequence.
ACCESSION  AA509926
VERSION    AA509926.1 GI:2247780
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
            Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
            Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
            Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
            Waterston, R.
            The WashU-HMI Mouse EST Project
            Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:506954
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 425.
            Location/Qualifiers
            1. 456
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:862866"
            /clone_lib="Soares_mammary_gland_NBMG"
            /sex="male"
            /tissue_type="mammary gland"
            /dev_stage="4 weeks"
            /lab_host="DH10B"
            /note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
            ) with a modified polylinker; Site_1: Not I; Site_2: Eco
            RI; 1st strand cDNA was primed with a Not I - oligo(dT)

```

TITLE

JOURNAL

COMMENT

FEATURES

source

alignment_scores:		
Quality:	8.00	Length: 8
Ratio:	1.000	Gaps: 0
Percent Similarity:	100.000	Percent Identity: 100.000

Percent

```
alignment_scores:      Length: 8
                      Quality: 8.00
                      Gaps: 0
                      Ratio: 1.000
                      Percent Identity: 100.000
```

```

alignment_block:
US-09-674-779-2 x AA874458/rev ..
Align seg 1/1 to reverse of: AA874458 from: 1 to: 459

122 HisLeuTrpGlyGlnIleValPro 129
|||||
138 CATCTCTGGGGCAGATAGTACCA 115

seq_name: gb_est2:BF514220

seq_documentation_block:
LOCUS BF514220 464 bp mRNA linear EST 29-AUG-2001
DEFINITION BBI60014B10H08 Bee Brain Normalized Library, BBI6 Apis mellifera
CDNA clone BBI60014B10H08 5', mRNA sequence.
ACCESSION BF514220
VERSION BF514220.1 GI:15364594
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
REFERENCE 1 (bases 1 to 464)
AUTHORS Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
Smollier,D. and Robinson,G.E.
TITLE An Expressed Sequence Tag Resource for Studies of Brain and
Behavior in the Honey Bee
JOURNAL Unpublished (2001)
COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCACTAAG
Insert Length: 464 Std Error: 0.00
Plate: BBI60014B10 row: H column: 08
Seq primer: AGCGATACATTTTCACACGGA
High quality sequence stop: 464.
FEATURES
source Location/Qualifiers
1..464
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BBI60014B10H08"
/clone_lib="Bee Brain Normalized Library, BBI6"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT73-Pac; Site_1: EcoRI;
Site_2: NotI; The BBI6 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996). Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
BASE COUNT 155 a 73 c 99 g 137 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000
Percent Similarity: 100.000 Gaps: 0
Percent Identity: 100.000

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

seq_name: gb_est2:BF565248

seq_documentation_block:
LOCUS BF565248 465 bp mRNA linear EST 12-DEC-2000
DEFINITION UI-R-B01-ajj-c-10-0-UI.r1 UI-R-B01 Rattus norvegicus cDNA clone
UI-R-B01-ajj-c-10-0-UI 5', mRNA sequence.
ACCESSION BF565248
VERSION BF565248.1 GI:11674978
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 465)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID- 1798296
Seq primer: M13 Forward
FEATURES
source Location/Qualifiers
1..465
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-B01-ajj-c-10-0-UI"
/clone_lib="UI-R-B01"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The library
(UI-R-B01) is a subtracted library derived from a mixture
of the following tissues: thalamus, cerebellum,
hypothalamus, medulla, pons, midbrain, cerebral
corpus striatum and hippocampus. For a detailed
description of the library from which this clone was
derived, please visit our web site at
rategen.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon and Soares, Genome Research
6:791-806, 1996)"
BASE COUNT 109 a 148 c 147 g 60 t 1 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000
Percent Similarity: 100.000 Gaps: 0
Percent Identity: 100.000

```

Tue Sep 17 07:27:54 2002

```

alignment_block:
US-09-674-779-2 x BF565248/rev ..
Align seg 1/1 to reverse of: BF565248 from: 1 to: 465

101 LeuLeuThrAlaArgSerTrp 108
|||||
260 CTCTCAACACTGCGGTCATGG 237

seq_name: gb_est2:BI513342

seq_documentation_block:
LOCUS BI513342 477 bp mRNA linear EST 29-AUG-2001
DEFINITION BB160012B10B07 Bee Brain Normalized Library, BB16 Apis mellifera
cDNA clone BB160012B10B07 5', mRNA sequence.
ACCESSION BI513342
VERSION BI513342.1 GI:15363716
KEYWORDS EST.
SOURCE honeybee.
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata
; Apoidea; Apidae; Apis.
1 (bases 1 to 477)
Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
Smoller,D. and Robinson,G.E.
An Expressed Sequence Tag Resource for Studies of Brain and
Behavior in the Honey Bee
Unpublished (2001)
JOURNAL
COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATAGCATCTACTATGAGG
BACKWARD: ATTACCCCTCACTAAG
Insert Length: 477 Std Error: 0.00
Plate: BB160012B10 row: B column: 07
Seq primer: ACGGATAACAATTTCACACAGGA
High quality sequence stop: 477.
FEATURES
source
1..477
Location/Qualifiers
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7450"
/clone="BB160012B10B07"
/clone_lib="Bee Brain Normalized Library, BB16"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoRI;
Site_2: NotI. The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
BASE COUNT 159 a 76 c 100 g 142 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent similarity: 100.000 Percent Identity: 100.000
US-09-674-779-2 x BF565248/rev ..
Align seg 1/1 to reverse of: BF565248 from: 1 to: 465

```

```

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BI513342 ..
Align seg 1/1 to: BI513342 from: 1 to: 477

208 GlnAsnPheGlyLeuGlyLeuTyr 215
|||||
74 CAAAATTTGGTTGGGATTATAT 97

seq_name: gb_gss:BH105280

seq_documentation_block:
LOCUS BH105280 487 bp DNA linear GSS 19-JUL-2001
DEFINITION RPCI-24-351K23.TJ RPCI-24 Mus musculus genomic clone RPCI-24-351K23
, DNA sequence.
ACCESSION BH105280
VERSION BH105280.1 GI:14934069
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 487)
Zhao,S., Nieman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Other_GSSs: RPCI-24-351K23.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: shao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 351 row: K column: 23
Seq primer: SP6
Class: BAC ends.
FEATURES
source
1..487
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-351K23"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 125 a 103 c 141 g 118 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent similarity: 100.000 Percent Identity: 100.000
US-09-674-779-2 x BH105280/rev ..
Align seg 1/1 to reverse of: BH105280 from: 1 to: 487

```

169 LysHisLeuThrAsnSerAlaIle 176
 |||||
 260 AAGCACCTTACTAACTACGCCATC 237

seq_name: gb_est2:BI512001

seq_documentation_block:

LOCUS BI512001 489 bp mRNA linear EST 29-AUG-2001
 DEFINITION BB160007B20D07 Bee Brain Normalized Library, BB16 Apis mellifera
 cDNA clone BB160007B20D07 5', mRNA sequence.

ACCESSION BI512001
 VERSION BI512001.1 GI:15362375
 KEYWORDS EST.
 SOURCE honeybee.

ORGANISM

Apis mellifera
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Hymenoptera; Apoidea; Aculeata
 ; Apoidea; Apidae; Apis.

REFERENCE 1 (bases 1 to 489)

Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
 Smoller,D. and Robinson,G.E.

Behavior in the Honey Bee

Unpublished (2001)

CONTACT: Gene E. Robinson

Department of Entomology

University of Illinois

505 S. Goodwin Ave., Urbana, IL 61801, USA

Tel: 217 265 0309

Fax: 217 244 3499

Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical
 Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
 Award in Functional Genomics to G.E. Robinson and an NSF
 Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG

BACKWARD: ATTAACCTCTACTAAAG

Insert Length: 489 Std Error: 0.00

Plate: BB160007B20 row: D column: 07

Seq primer: AGCGGATACCAATTCACACAGGA

High quality sequence stop: 489.

Location/Qualifiers

1..489

/organism="Apis mellifera"

/strain="mixed strains of European bees, predominantly

A.m. ligustica"

/db_xref="taxon:7460"

/clone="BB160007B20D07"

/clone_lib="Bee Brain Normalized Library, BB16"

/sex="female"

/tissue_type="brain"

/dev_stage="adult worker honey bee"

/lab_host="DH10B"

/note="Organ: brain; Vector: pT73-Pac; Site:1: EcoRI;

Site:2: NotI; The BB16 library was contributed by the

Soares laboratory and it was constructed and normalized

as described by Bonaldo, M.F., Lennon, G. and Soares,

M.B. (1996), Genome Research 6(9): 791-806. RNA was

prepared from dissected brains of adult worker bees of

various ages and various behavioral groups."

BASE COUNT 164 a 79 c 105 g 141 t

ORIGIN

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Length: 8

Gaps: 0

Percent Identity: 100.000

US-09-674-779-2 x BI512001

alignment_block:

Align seg 1/1 to: BI512001 from: 1 to: 489

208 GlnAsnPhcGlyLeuGlyLeuTyr 215

|||||
 67 CAAATTTTGGTTGGGATTATAT 90

seq_name: gb_est2:BI515132

seq_documentation_block:

LOCUS BI515132 516 bp mRNA linear EST 29-AUG-2001

DEFINITION BB160017B10C08 Bee Brain Normalized Library, BB16 Apis mellifera

cDNA clone BB160017B10C08 5', mRNA sequence.

ACCESSION BI515132

VERSION BI515132.1 GI:15365506

KEYWORDS EST.

SOURCE honeybee.

ORGANISM

Apis mellifera

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Hymenoptera; Apoidea; Aculeata

; Apoidea; Apidae; Apis.

REFERENCE 1 (bases 1 to 516)

Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,

Smoller,D. and Robinson,G.E.

An Expressed Sequence Tag Resource for Studies of Brain and

Behavior in the Honey Bee

Unpublished (2001)

CONTACT: Gene E. Robinson

Department of Entomology

University of Illinois

505 S. Goodwin Ave., Urbana, IL 61801, USA

Tel: 217 265 0309

Fax: 217 244 3499

Email: generobi@life.uiuc.edu

This research was funded by the University of Illinois Critical

Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation

Award in Functional Genomics to G.E. Robinson and an NSF

Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.

PCR Primers

FORWARD: TAATACGACTCACTATAGG

BACKWARD: ATTAACCTCTACTAAAG

Insert Length: 516 Std Error: 0.00

Plate: BB160017B10 row: C column: 08

Seq primer: AGCGGATACCAATTCACACAGGA

High quality sequence stop: 516.

Location/Qualifiers

1..516

/organism="Apis mellifera"

/strain="mixed strains of European bees, predominantly

A.m. ligustica"

/db_xref="taxon:7460"

/clone="BB160017B10C08"

/clone_lib="Bee Brain Normalized Library, BB16"

/sex="female"

/tissue_type="brain"

/dev_stage="adult worker honey bee"

/lab_host="DH10B"

/note="Organ: brain; Vector: pT73-Pac; Site:1: EcoRI;

Site:2: NotI; The BB16 library was contributed by the

Soares laboratory and it was constructed and normalized

as described by Bonaldo, M.F., Lennon, G. and Soares,

M.B. (1996), Genome Research 6(9): 791-806. RNA was

prepared from dissected brains of adult worker bees of

various ages and various behavioral groups."

BASE COUNT 152 a 100 c 122 g 142 t

ORIGIN

alignment_scores:

Quality: 8.00

Ratio: 1.000

Percent Similarity: 100.000

Length: 8

Gaps: 0

Percent Identity: 100.000

Tue Sep 17 07:27:54 2002

```

alignment_block:
US-09-674-779-2 x B1515132 ..
Align seg 1/1 to: B1515132 from: 1 to: 516

208 GlnAsnPheGlyLeuGlyLeuTyR 215
|||||
75 CAGAAATTCGGATTAGGCCTTAT 98

seq_name: gb_gss:B48301

seq_documentation_block:
LOCUS B48301 528 bp DNA linear GSS: 08-APR-1999
DEFINITION RPCI11-6P4.TV RPCI-11 Homo sapiens genomic clone RPCI-11-6P4, DNA
sequence.
ACCESSION B48301
VERSION B48301.1 GI:2600538
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 528)
Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter
,J.C.
Use of BAC End Sequences for Sequence-Ready Map Building
Unpublished (1997)
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamadset@ig.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piet@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES
Location/Qualifiers
source
1..528
/organism="Homo sapiens"
/db_xref="GDB:7502283"
/db_xref="taxon:9606"
/clone="RPCI-11-6P4"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pRACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 153 a 114 c 76 g 165 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x B48301 ..
Align seg 1/1 to: B48301 from: 1 to: 528

240 IleCysArgHisValLeuProLys 247
|||||
11 ATCTGTAGACATGATTACCAAAA 34

seq_name: gb_est2:B1643242

seq_documentation_block:
LOCUS B1643242 535 bp mRNA linear EST 29-AUG-2001
DEFINITION RS4_C01 Sugar beet root cDNA library (subtracted) Beta vulgaris
cDNA 5', mRNA sequence.
ACCESSION B1643242
VERSION B1643242.1 GI:15545452
KEYWORDS EST.
SOURCE Beta vulgaris.
ORGANISM Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Caryophyllales; Chenopodiaceae; Beta.
1 (bases 1 to 533)
de los Reyes,B.G., McGrath,J.M. and Myers,S.
Differential gene expression in sugar beet (beta vulgaris)
Unpublished (2001)
Contact: J. Mitchell McGrath
Sugar Beet Genetics Lab., USDA/ARS/Sugar Beet and Bean Research
Unit
Michigan State University
494 Plant and Soil Sci. Bldg., East Lansing MI 48824, USA
Tel: (517)-432-2355
Fax: (517)-337-6782
Email: mitchmcg@pilot.msu.edu
Seq primer: T3.
FEATURES
Location/Qualifiers
source
1..533
/organism="Beta vulgaris"
/cultivar="USH20"
/db_xref="taxon:161934"
/clone_lib="Sugar beet root cDNA library (subtracted)"
/tissue_type="mature root"
/lab_host="XLOLR"
/note="Organ: root; Vector: pBK-CMV; Site_1: EcoRI;
Site_2: XhoI; cDNAs were derived from reverse
transcription of mRNA samples from mature root tissue. The
mRNA pool was used as template for double stranded cDNA
synthesis using the Stratagene pBluescript XR cDNA
synthesis and library kit. The resulting cDNA was used as
a tester for subtraction against a driver cDNA population
derived from leaf, inflorescence, and 4-day old germinated
seedlings. Subtraction was performed using the Invitrogen
photobiotin-Streptavidin subtractor kit. The cDNA library
was generated by directional ligation of the subtracted
cDNAs in the EcoRI and XhoI sites of pBK-CMV vector
(Stratagene). Putative ID reported when E scores were
better than e-10 via nr database and blastx."
BASE COUNT 193 a 79 c 130 g 131 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x B1643242/rev ..
Align seg 1/1 to reverse of: B1643242 from: 1 to: 533

89 SerArgLeuGlyAsnTyrLeuPro 96
|||||
174 AGTCGCTTAGGAACACTACTTGCCA 151

seq_name: gb_est2:B1517071

seq_documentation_block:
LOCUS B1517071 535 bp mRNA linear EST 29-AUG-2001
DEFINITION BB160024A20E01 Bee Brain Normalized Library, BB16 Apis mellifera
cDNA clone BB160024A20E01 5', mRNA sequence.
ACCESSION B1517071

```

```

VERSION      BI517071.1  GI:15367445
KEYWORDS     EST.
SOURCE       honeybee.
ORGANISM     Apis mellifera
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Hymenoptera; Apoecrita; Aculeata
; Apoidea; Apidae; Apis.
REFERENCE    1 (bases 1 to 535)
AUTHORS      Whitfield,C.W., Soares,B., Robertson,H.M., Pardinas,J., Liu,L.,
Smoller,D. and Robinson,G.E.
TITLE        An Expressed Sequence Tag Resource for Studies of Brain and
Behavior in the Honey Bee
JOURNAL      Unpublished (2001)
COMMENT      Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR Primers
FORWARD: TAATAGACTCTACTATAGGG
BACKWARD: ATTAACCTCTACTAAAG
Insert Length: 535 Std Error: 0.00
Plate: BB160024A20 row: E column: 01
Seq primer: AGCGGATACAAATTCACACAGGA
High quality sequence stop: 535.
FEATURES     source
source       1..535
/organism="Apis mellifera"
/strain="mixed strains of European bees, predominantly
A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB160024A20E01"
/clone_lib="Bee Brain Normalized Library, BB16"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/notes="Organ: brain; Vector: p7T3-Pac; Site_1: EcoRI;
Site_2: NotI; The BB16 library was contributed by the
Soares laboratory and it was constructed and normalized
as described by Bonaldo, M.F., Lennon, G. and Soares,
M.B. (1996), Genome Research 6(9): 791-806. RNA was
prepared from dissected brains of adult worker bees of
various ages and various behavioral groups."
BASE COUNT   183 a 103 c 121 g 128 t
ORIGIN

alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BI517071 ..
Align seg 1/1 to BI517071 from: 1 to: 535
208 GlnAsnPheGlyLeuGlyLeuTyr 215
|||||
17 CAGAAATTCGATAGCCCTTTAT 40

seq_name: gb_estl:AI995165
seq_documentation_block:
LOCUS      AI995165
DEFINITION 701502426 A. thaliana, Ohio State clone set Arabidopsis thaliana

alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x BI517071 ..
Align seg 1/1 to BI517071 from: 1 to: 535
208 GlnAsnPheGlyLeuGlyLeuTyr 215
|||||
17 CAGAAATTCGATAGCCCTTTAT 40

seq_name: gb_estl:AI995165
seq_documentation_block:
LOCUS      AI995165
DEFINITION 701502426 A. thaliana, Ohio State clone set Arabidopsis thaliana

alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AI995165/rev ..
Align seg 1/1 to reverse of: AI995165 from: 1 to: 544
144 ProAlaAsnThrGlnIleArgSer 151
|||||
483 CCGGCCAATACACAGATCCGATCC 460

seq_name: gb_gss:AQ266185
seq_documentation_block:
LOCUS      AQ266185
DEFINITION CITBI-E1-2502L22.TF CITBI-E1 Homo sapiens genomic clone 2502L22,
DNA sequence.
ACCESSION  AQ266185
VERSION     AQ266185
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 579)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE       Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL      Unpublished (1998)
COMMENT      Contact: Mark Adams
Department of Eukaryotic Genomics

CDNA clone 701502426, mRNA sequence.
ACCESSION    AI995165
VERSION      AI995165.1  GI:5842070
KEYWORDS     EST.
SOURCE       thale cress.
ORGANISM     Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 544)
Chen,J., Hillman,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
Wang,X., Hilleman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.,
Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobruga,A., Murry,L.,
Turner,C., Krikorian,S., Elder,L. and Hanson,D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)
Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.
Location/Qualifiers
1..544
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone="701502426"
/clone_lib="A. thaliana, Ohio State clone set"
/notes="cDNA library was made from selected clones from the
Arabidopsis thaliana Ohio State clone set."
BASE COUNT   165 a 98 c 141 g 140 t
ORIGIN

alignment_scores:
Quality:      8.00      Length:      8
Ratio:        1.000     Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AI995165/rev ..
Align seg 1/1 to reverse of: AI995165 from: 1 to: 544
144 ProAlaAsnThrGlnIleArgSer 151
|||||
483 CCGGCCAATACACAGATCCGATCC 460

seq_name: gb_gss:AQ266185
seq_documentation_block:
LOCUS      AQ266185
DEFINITION CITBI-E1-2502L22.TF CITBI-E1 Homo sapiens genomic clone 2502L22,
DNA sequence.
ACCESSION  AQ266185
VERSION     AQ266185
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 579)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE       Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL      Unpublished (1998)
COMMENT      Contact: Mark Adams
Department of Eukaryotic Genomics

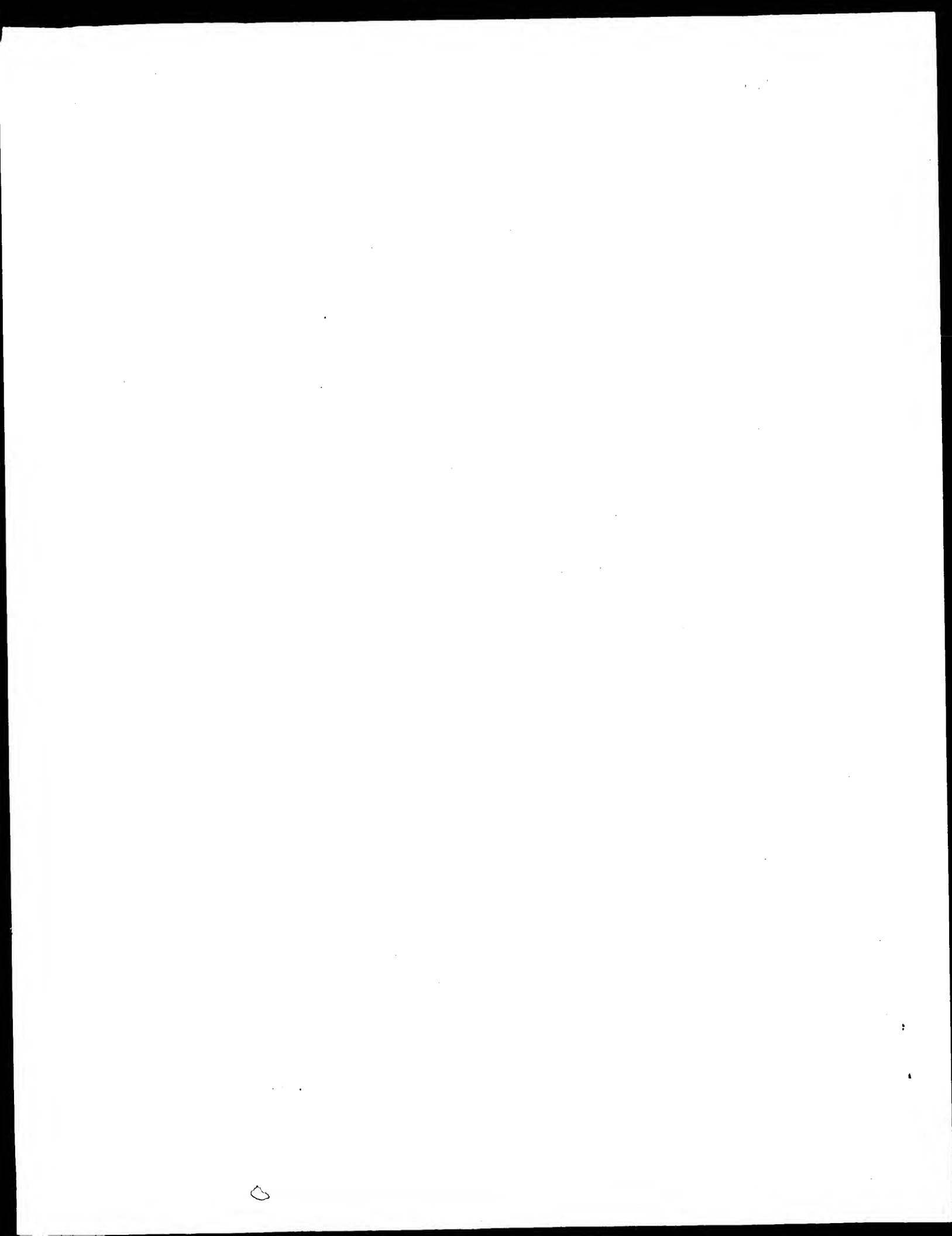
```

PCR Primers
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCCCTCACTAAAG
Insert Length: 580 Std Error: 0.00

Tue Sep 17 07:27:54 2002

us-09-674-779-2.oli.rst

Page 23



ON of: US-09-674-779-2.g: GenEmbl:* , out_format : pfs
Date: Sep 16, 2002 6:52 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+P2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09674779/tunat_12092002.124121_25832/app_query.fasta_1.310
-DB=GenEmbl -QFWT=fastap -SUFFIX=std.rge -GAPOP=12.000
-GAPEXT=4.000 -MINWATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPOP=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFWT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779-EGN1_1.4034 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-2
Query length: 250
Database: GenEmbl:*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 1868.280000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AX081145	+ 1355.00	2163.56	3.8e-112	753	AX081145 Sequence 1 from Patent
gb_pat:AX067442	+ 1355.00	2131.81	2.2e-110	23210	AX067442 Sequence 17 from Patent
gb_ba:AF000582	+ 375.00	571.09	1.9e-23	8367	AF000582 Acinetobacter sp. ADP
gb_ba:SM591791	+ 127.00	139.41	21.10	340300	AL591791 Sinorhizobium meli
gb_ba:AP003009	+ 112.00	115.41	458.27	339681	AP003009 Mesorhizobium loti
gb_hg:AC046147	+ 110.00	121.81	201.66	120405	AC046147 Mus musculus chrom
gb_hg:AC073821	+ 110.00	120.07	252.19	145347	AC073821 Mus musculus clone
gb_hg:AC073767	+ 110.00	117.05	371.47	201377	AC073767 Mus musculus clone
gb_ba:AP003000	+ 106.50	106.37	1.5e+03	348077	AP003000 Mesorhizobium loti
gb_pr:AC068614	+ 106.00	112.50	665.59	164706	AC068614 Homo sapiens BAC cl
gb_hg:AC098619	+ 105.50	110.67	842.27	184172	AC098619 Rattus norvegicus c
gb_hg:AC030643	+ 103.50	108.62	1.1e+03	162527	AC030643 Homo sapiens chrom
gb_hg:AC027289	+ 103.50	108.00	1.2e+03	173797	AC027289 Homo sapiens chrom
gb_ba:AE005826	+ 103.00	131.55	57.82	12528	AE005826 Caulobacter crescent
gb_hg:AC003656_4	+ 102.00	109.83	937.46	110000	Continuation (5 of 8) of AC0
gb_pr:AP001053	+ 102.00	106.25	1.5e+03	161920	AP001053 Homo sapiens genom
gb_hg:AC079571	+ 102.00	102.41	2.4e+03	245130	AC079571 Mus musculus clone
gb_pr:AP001752	+ 102.00	99.38	3.6e+03	340000	AP001752 Homo sapiens genom
gb_pr:AC078925	+ 101.50	105.96	1.5e+03	153284	AC078925 Homo sapiens 12q BA
gb_pr:AL133414	+ 101.50	105.61	1.6e+03	159169	AL133414 Human DNA sequence
gb_ba:AC065304	+ 101.50	105.49	1.6e+03	161247	AC065304 Homo sapiens BAC cl
gb_ba:AE008307	+ 101.00	128.89	81.34	11814	AE008307 Agrobacterium tumefa
gb_pr:AC009307	+ 101.00	128.81	82.15	11913	AE009307 Agrobacterium tumefa
gb_pr:AC106530	+ 100.50	103.06	2.2e+03	176186	AC106530 Homo sapiens chrom
gb_in:CAE223616	+ 100.50	102.15	2.5e+03	194508	AC106530 Rattus norvegicus c
gb_pr:HS8337	+ 100.00	136.47	30.77	4383	AC23616 Caenorhabditis elegan
gb_pr:AC083964	+ 100.00	108.84	1.1e+03	86574	AL008637 Human DNA sequence
gb_hg:AC090135	+ 100.00	103.09	2.2e+03	161160	AC083964 Homo sapiens chrom
gb_ba:AP003008	+ 100.00	101.56	2.7e+03	190173	AC090135 Homo sapiens chrom
gb_hg:AC105039	+ 100.00	95.99	5.5e+03	346820	AP003008 Mesorhizobium loti
gb_hg:AC016944	+ 99.50	103.22	2.2e+03	145667	AC105039 Homo sapiens chrom
gb_pr:AL139038	+ 99.50	102.42	2.4e+03	158886	AC016944 Homo sapiens chrom
gb_hg:AC073525	+ 99.00	102.74	2.3e+03	140756	AL139038 Human DNA sequence
gb_hg:AC103346	+ 99.00	99.67	3.4e+03	196048	AC073525 Homo sapiens chrom
gb_in:AC001394	+ 98.50	116.24	412.26	30062	AC103346 Rattus norvegicus cl
gb_in:AC009905	+ 98.50	100.22	3.2e+03	169516	AC013934 Drosophila melanoga
gb_in:AC010121	+ 98.50	100.16	3.2e+03	170662	AC009905 Drosophila melanoga
gb_hg:AC087566	+ 98.50	100.05	3.3e+03	172529	AC010121 Drosophila melanoga
gb_hg:AC087566	+ 98.50	97.32	4.7e+03	231912	AC087566 Mus musculus clone
gb_in:AE003655	+ 98.50	96.48	5.2e+03	253753	AE003655 Drosophila melanoga

gb_pr:HS278N12 + 98.00 106.54 1.4e+03 78533 ! AL033376 Human DNA sequenc
gb_hg:AC097543 + 98.00 100.78 3.0e+03 146321 ! AC097543 Rattus norvegicu
gb_in:PMH1SH2A + 97.50 148.33 6.72 790 ! X01346 Psammochinus miliaris
gb_in:SUPHISB4 + 97.50 147.42 7.56 872 ! M10559 P. miliaris histone 4?
em_hggo_inv:AC034780 + 97.50 145.56 9.59 1066 ! AC034780 Giardia intestinal
seq_name: gb_pat:AX081145
seq_documentation_block:
LOCUS AX081145 753 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 1 from Patent WO0109335.
ACCESSION AX081145
VERSION AX081145.1 GI:13170040
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
REFERENCE 1 (bases 1 to 753)
AUTHORS Thompson, J.
TITLE Moraxella catarrhalis antigen basb120
JOURNAL Patent: WO 0109335-A 1 08-FEB-2001;
SmithKline Beecham Biologicals S.A. (BE)
FEATURES
source
i. 753
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 239 a 166 c 152 g 196 t
ORIGIN

alignment_scores:

Quality: 1355.00 Length: 250
Ratio: 5.420 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AX081145

Align seg 1/1 to: AX081145 from: 1 to: 753

1 MetLysAsnPhenGlnTyrPheIleThrThreuleSerMetLe 17
|||||
1 ATGAAATTTTAACTAACTTATTAACATACATTTATCAGCATGCT 50
17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
|||||
51 GGTTGCATGCAGTGCACCCATACCAACCAATCCCAAGTATCCCAATAA 100
34 yThrProSerValLeuIleThrLysAspLysIleGlyAspHisThr 50
|||||
101 AAACGCCATCGTACTGATTAAGTAAAGATAAAATCGGTGATCATCATACA 150
51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG 67
|||||
151 CATGACGACATGAATCTGTAAGCCATGCGGTTCGAGGCACATTTTGA 200
67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
|||||
201 GACTTGGCTACAGATGCACCATGCCACCAACACAGAGGTAGTAGTATC 250
84 InAlaTrpLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
|||||
251 AGCGGATTTTACAATCAAGACTTGGTAAATATATCGCCACCAATGAGTCAA 300
101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG 117
|||||
301 CTACTAACACATCGCAGCGTCATGGCAGGATGTCATGACACCTTATCA 350
117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeu 134
|||||
351 GCTGCCACCAGAACATCTTTGGGGTCAGATGTTGACCAACATTCACCTGT 400

134 yrGlnAspLeuLysSerArgGlyLeuLeuProAlaAsnThrGlnIleArg 150
11583 AAAGCCATCGGTACTGATTAATAAGATATAATCGGTGATCATATACA 11632
11633 CATGAGCAGCATGAATCTGTAAAGCCATGTCGGTTTGCAGGCACATTTGA 11682
51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
11683 GACTGGCTACAGATGCACCATGCGCCACCAACAAGAGGTAGTAGTATC 11732
84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGln 100
11733 AGCGTATTACAAATCAAGACTTGGTAATATATCTGCCACCAATGAGTCAA 11782
101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG1 117
11783 CTACTAAGCACTGCACGGTCATGGCAGGCATGGTCATGACCTTATCA 11832
117 nLeuProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
11833 CHTGCCACAGACATCTTTGGGGTCAGATTGTACCAACATTCGACTTGT 11882
134 yrGlnAspLeuLysSerArgGlyLeuLeuProAlaAsnThrGlnIleArg 150
11883 ATCAAGATCTCAAAAGTAGGGGCATATGCGCAGCAATACCCAAATTCGC 11932
151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
11933 TCAGTTTATCGCAATCCTGAACCTCAACCAATGTGCTGGTGCAGCTAT 11982
167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
11983 GAGTAAGCATTTGACCAATAGTCCATTTGATTTGGGTGGCTGACCTTG 12032
184 luileLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
12033 AAATAAAAGCCAGGCACCTGTATGAGCTTCAAAACCGCTTATGCCAATAT 12082
201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
12083 TGGGTAGCATGCGCAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 12132
217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
12133 AGGGCGATTCTCTGGATACCCAGGGTTAGAAATGGGTGCTCAAT 12182
234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
12183 TTTCTGAAACAAACTCTATTGTCGTCATGCTTTACCAAAAAATAAGCTA 12232

seq_name: gb_pat:AX067442

seq_documentation_block:
LOCUS AX067442 23210 bp DNA linear PAT 24-JAN-2001

DEFINITION Sequence 17 from Patent WO0078968.

ACCESSION AX067442

VERSION AX067442.1 GI:12545062

KEYWORDS Moraxella catarrhalis.

SOURCE Moraxella catarrhalis.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 23210)

AUTHORS Lagace, R.E., Patterson, C. and Berg, K.L.

TITLE Nucleotide sequences of moraxella catarrhalis genome

JOURNAL Patent: WO 0078968-A 1, 28-DEC-2000;

FEATURES Location/Qualifiers

source 1..23210

/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

unsure

/note="a or g or c or t, unknown, or other"

BASE COUNT 7108 a 4482 c 4951 g 6668 t 1 others

ORIGIN

alignment_scores:

Quality: 1355.00 Length: 250

Ratio: 5.420 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AX067442

Align seg 1/1 to: AX067442 from: 1 to: 23210

1 MetLysAsnPheAsnGlnTyrPheIleThrThrIleSerSerMetLeu 17

11583 AAAGCCATCGGTACTGATTAATAAGATATAATCGGTGATCATATACA 11632
11633 CATGAGCAGCATGAATCTGTAAAGCCATGTCGGTTTGCAGGCACATTTGA 11682
51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
11683 GACTGGCTACAGATGCACCATGCGCCACCAACAAGAGGTAGTAGTATC 11732
84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGln 100
11733 AGCGTATTACAAATCAAGACTTGGTAATATATCTGCCACCAATGAGTCAA 11782
101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG1 117
11783 CTACTAAGCACTGCACGGTCATGGCAGGCATGGTCATGACCTTATCA 11832
117 nLeuProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
11833 CHTGCCACAGACATCTTTGGGGTCAGATTGTACCAACATTCGACTTGT 11882
134 yrGlnAspLeuLysSerArgGlyLeuLeuProAlaAsnThrGlnIleArg 150
11883 ATCAAGATCTCAAAAGTAGGGGCATATGCGCAGCAATACCCAAATTCGC 11932
151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
11933 TCAGTTTATCGCAATCCTGAACCTCAACCAATGTGCTGGTGCAGCTAT 11982
167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
11983 GAGTAAGCATTTGACCAATAGTCCATTTGATTTGGGTGGCTGACCTTG 12032
184 luileLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
12033 AAATAAAAGCCAGGCACCTGTATGAGCTTCAAAACCGCTTATGCCAATAT 12082
201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
12083 TGGGTAGCATGCGCAAAACCAAAATTTTGGGCTGGGTTTATACGCCAC 12132
217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
12133 AGGGCGATTCTCTGGATACCCAGGGTTAGAAATGGGTGCTCAAT 12182
234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
12183 TTTCTGAAACAAACTCTATTGTCGTCATGCTTTACCAAAAAATAAGCTA 12232

seq_name: gb_pat:AX067442

seq_documentation_block:

LOCUS AF400582 8367 bp DNA linear BCT 08-NOV-2001

DEFINITION Acinetobacter sp. ADP1 mismatch repair protein (mutS), 7-Fe

ferredoxin (fdxA), and O-methyltransferase-like protein genes,

complete cds; and unknown genes.

ACCESSION AF400582

VERSION AF400582.1 GI:15217080

KEYWORDS Acinetobacter sp. ADP1.

SOURCE Acinetobacter sp. ADP1.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 8367)

AUTHORS Young, D.M. and Ormston, L.N.

TITLE Functions of the Mismatch Repair Gene mutS from Acinetobacter sp.

JOURNAL J. Bacteriol. 183 (23), 6822-6831 (2001)

MEDLINE 21555098

PUBMED 11698371

seq_name: gb_ba:SME591791	seq_documentation_block:	340900 bp	DNA	linear	BCT 16-AUG-2001
LOCUS	SME591791	340900 bp	DNA	linear	BCT 16-AUG-2001
DEFINITION	Sinorhizobium meliloti 1021 complete chromosome; segment 10/12.				
ACCESSION	AL591791	AL591688			
VERSION	AL591791.1	GI:15075538			
KEYWORDS					
SOURCE	Sinorhizobium meliloti.				
ORGANISM	Sinorhizobium meliloti				
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
AUTHORS	1 Rhizobiaceae; Sinorhizobium.				
	1 (bases 1 to 340900)				
	Capella,D., Barclay-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,				
	Bolstad,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,				
	Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,				
	Pohl,T., Portetelle,D., Puhler,A., Purnelle,B., Ramsperger,U.,				
	Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.				
	From the Cover: Analysis of the chromosome sequence of the legume				
	symbiont Sinorhizobium meliloti strain 1021				
TITLE	proceedings of the National Academy of Sciences of the United				
JOURNAL	States of America. 98 (17), 9877-9882 (2001)				
	11481430				
PUBMED	2 (bases 1 to 340900)				
REFERENCE	Gouzy,J.				
AUTHORS	Direct Submission				
TITLE	Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELIOL				
JOURNAL	EU Consortium				
COMMENT	MELIOL EU Consortium:				
	Laboratoire de Biologie Molculaire des Relations				
	Plantes-Microorganismes, UMR215-CNRS-INRA, BP21, F-31326 Castanet,				
	France, Laboratoire de Genetique et developpement UMR6061-CNRS,				
	Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,				
	France. GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,				

```

predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE ALTERNATIVE CYTOCHROME C OXIDASE
POLYPEPTIDE I TRANSMEMBRANE PROTEIN"
/protein_id="CAC47096.1"
/db_xref="GI:15075541"
/translation="MWDVRSRGGEALPPPEVEDVELYHPHSWTRVYFVSQDAKIIAI
QYSTAIAIGMVALVLSLRLQLOFGFTFELIDAERYQFIMHGIMVYILITALE
LGGFNTLIPLMGARDMVPYANMLSYWILLAVIVLAASFPTPGPTGAGWTLVPP
QAVLSGTPGGRDGMIMMLSLIIFVIGFTMGGLNVVTVVLOGRAGMTLMRLPLTVM
GIFTATWALLAFALFVACVMDLFDRLCTSFEMPAIVEMGEOLQYGGSPILFOHL
FWRFGEVIVALPAEGLIVDLSISHARNKIEGYRMWVAIVLIGLSFTVWAHMY
VSGNPPYEGFEFATTLIIAAPTAKYNNVLTRGNHILTLPLMLFALAFIVTFYNG
GLIGFIGNVVVDPLSDTMFVVAHFMVNGVAPIMVIFGAIYHWPKITGRMLNEAM
QGHFWFTFGAIAFFPMHYLGLIGVPRRYHELGEASFVTTSTIAELNAPISVMALLV
GAAQIVFLNLAHLRGAGGNPWRATLEWTPTETPPPHGNWGNELPVVYRWAYD
YSVGEAPEDFIPQNPPTGRLSHETVS"
3412..4110
/gene="coxO OR SMC01984"
CDS
3412..4110
/gene="coxO OR SMC01984"
/EC_number="1.9.3.-"
/function="small molecule metabolism; energy transfer;
electron transport"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE CYTOCHROME-C OXIDASE TRANSMEMBRANE
PROTEIN"
/protein_id="CAC47097.1"
/db_xref="GI:15075542"
/translation="MSIAFFLAIAIITAWLAGORLTSRPVLEVGHPHDRGATRL
PPAKIGLVFLAVVGLFSLAISAYFWRNASDWDGALPLGLLNTMTGLAAGSITLH
WTKVEARRNDEARGLIAGLALGALFAGQLFAWRALSDAGYFLAGNPANSFYLL
TGMGHGIIIGGLFALGRVTAHASQTPGLGNRTLSELCAIYWHFLLVVLVLFALFAG
WASGVIEFCRQLLT"
4123..4845
/gene="coxP OR SMC01985"
CDS
4123..4845
/gene="coxP OR SMC01985"
/EC_number="1.9.3.-"
/function="small molecule metabolism; energy transfer;
electron transport"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="PROBABLE CYTOCHROME-C OXIDASE PROTEIN"
/protein_id="CAC47098.1"
/db_xref="GI:15075543"
/translation="NSAPMKPAAETLPRPAGLAGLADWASDORTFKDVSNGKMMW
IFLLSDTFVGCFLAYMSARMTSVPPNPSEVPALEIGGTHPLILIAIMTFVLIS
SSGTWAMVNYGYRRDRKRTAALMLLTALFGAAFVGMQAFESKLIAGVVRPWNPGW
AAQFGSTFMITGFHGTGTFVIFLLIVARKVWKGDFETERRRGGFTTSRKGRYEIVEI
TGLYWHFVDLVWVFIFAFFYLM"
4857..5216
/gene="SMC01986"
CDS
4857..5216
/gene="SMC01986"
/function="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical

```

```

predicted by Codon_usage
predicted by Framed"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="HYPOTHEICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAC47099.1"
/db_xref="GI:15075544"
/translation="WAHRETHPAQHAHAATEHQOHAIKYLLVWGLLFLVLSALSYLVD
YFLOGYLEWSLILIFMMLKAGLIVAVFMHMAWERLALIYILPLLLVLFVFLMWS
ESNYLFLRLTFEGGE"
5477..8038
/gene="SMC01987"
CDS
5477..8038
/gene="SMC01987"
/function="small molecule metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/transl_table=11

```

```

alignment_scores:
  Quality: 127.00      Length: 230
  Ratio: 1.187        Gaps: 13
  Percent Similarity: 46.522  Percent Identity: 23.913
alignment_block:
US-09-674-779-2 x SME591791/rev ..

```

```

Align seg 1/1 to reverse of: SME591791 from: 1 to: 340900
20 CysSerAlaProLeuProThrAsnProGln..... 29
299697 TGTCTGGCGGTCTGCTGAAAGGCCGCCAGGTCTCTGGCGTCCATCGCGT 299648
30 ValSerProIleuLysThrProSerValLeuLleuThrLysAspLysIleG 46
299647 GGTTCGTCTCTGTACACCCCGCATGGCGCTCCGCGTCCGCGCCGCG 299598
46 LysPheHisThrHisGluHisAspLysValSerHisValGlyLeu 62
299597 GCCAGACACGCACC.....CTC 299581
63 GlnAlaHisPheGluThrTrpLeuGlnMetHisAlaThrLysGlnG 79
299580 AACCTCTATTTC.....ATCCACACCAAGAAAGCGCA 299546
79 uValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr...L 95
299545 GATCACC.....TACAAGCGCAACGCTCGCTAGGATC 299514
95 euProMetSerGlnLeuLeuThrAlaArgSerTrpGlnAlaCys 111
299513 AGAAGGGTCTGCAGCAGATCAATCTCTCTCGCGACTGCGCGCA... 299467
112 GlyHisGluProTyrGlnLeuProGluHisLeu.....TrpG 125
299466 ...ACGACCCGACCAAAATGACCGCGCGCTCGATCTCGTCTGG... 299422
125 yGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyI 142
299421 .....GAAGTTTACCAGAGAGCGGTCTCCGGACT 299391
142 leuLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu 158
299390 ACATC.....CATGTCGTTTCCGCTATCGCTCCCGCCGACG 299353
159 Asn.....GlnCysAlaGlyGlyAlaAlaMetSerLysHis 170
299352 AACGGCATGCTGAGTCGCGGTGGAAGGCGTCTCGGAAGAGAGCCAGCA 299303

```

170 sleuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleYss 187
 187 erGlnAlaLeuTyTrGluLeuGlnAsnArgLeuCysGlnTyTrpLeuGlu 203
 299302 CATGTCGCAAGGCGTACTTATATACCGACGACGTGAAGCTCAAGA 299253
 299252 CG.....CTGCGCGAGTGGCGATGAAGTTC..... 299227
 204 HisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyAlaThrGlyAla.. 219
 299226CAGTGGCGGCGTGGCTATATCCGACTTCGGTTC 299189
 220leHisLeuAspThrGlnGlyPheArgLysTrp 230
 299188 GCCCTTCGTCATATGAGCTGGTGGCGTGGCGAGCTGG 299149
 seq_name: gb_ba:AP003009
 seq_documentation_block:
 LOCUS AP003009 339681 bp DNA linear BCT 15-MAY-2001
 DEFINITION Mesorhizobium loti DNA, complete genome, section 16/21.
 ACCESSION AP003009 BA000012
 VERSION AP003009.2 GI:14026063
 KEYWORDS
 SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.
 ORGANISM
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Phyllobacteriaceae; Mesorhizobium.
 1 (sites)
 Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
 Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
 Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
 Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,
 Takeuchi, C., Yamada, M. and Tabata, S.
 Complete genome structure of the nitrogen-fixing symbiotic
 bacterium Mesorhizobium loti
 DNA Res. 7 (6), 331-338 (2000)
 2 (bases 1 to 339681)
 Kaneko, T.
 Direct Submission
 Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
 Institute, The First Laboratory for Plant Gene Research; Yana
 1532-3, Kisarazu, Chiba 292-0812, Japan
 (E-mail:kaneko@kazusa.or.jp,
 URL:http://www.kazusa.or.jp/rhizobase/
 Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
 On May 11, 2001 this sequence version replaced gi:11994984.
 COMMENT
 FEATURES
 source
 1..339681
 /organism="Mesorhizobium loti"
 /strain="MAFF303099"
 /db_xref="taxon:381"
 complement(82..420)
 /gene="ml16350"
 complement(82..420)
 /gene="ml16350"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB52662.1"
 /db_xref="GI:14026064"
 /translation="MPLAFDLARIALPKVGLPIERNLEACSLTAWPHDGLRAEQ
 PVHKSQCRLWRSLLVHRHVLNEMRSIENNVRTILREACINWARGREPPVVCRS
 WPALTPCWK"
 447..593
 /gene="msr6351"
 447..593
 /gene="msr6351"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11

/protein_id="BAB52663.1"
 /db_xref="GI:14026065"
 /translation="MHALVHAADIQDCDGGVMVMAKLFGLDPLFLKLYAGGHSRGRS
 SKRG"
 complement(733..1464)
 /gene="ml16352"
 complement(733..1464)
 /gene="ml16352"
 /codon_start=1
 /transl_table=11
 /product="transposase"
 /protein_id="BAB52664.1"
 /db_xref="GI:14026066"
 /translation="MRTWAMSEVARDPLYRCHRPAAETIIAHVWLYFRFLSLRMVED
 MAAAGLIIVHOTIRSWAEKGRHFAREIKRRSAGCLGDKWHLDECYVAINGKQWLM
 RAYDQGLIVLYQSRNKAALKMLKLLKAGRTPRYMITDKLRSYDAARRDLMP
 GVEHRSHKLNNAENSHOPTRRRTMKRFSARQLORFVSIHDPANLHFPRNLT
 SSAQRDLRNAAQIWNKIAACAAA"
 complement(1080..1748)
 /gene="ml16353"
 complement(1080..1748)
 /gene="ml16353"
 /codon_start=1
 /transl_table=11
 /product="transposase"
 /protein_id="BAB52665.1"
 /db_xref="GI:14026067"
 /translation="MVEACATYKNNRFPETIVAHVWLYRFLSLRDEEMLELGERGI
 VVSETIRWCKKSPDYARLKRKAPSKDDVHLDVAGHRKVNRLMKMCEHGPOQ
 KLLIRIVATAFPPRSILRTYRGSTFAFRSACALRTCWRPAASSRRIRPEAGRRNS
 GGISPRSDGQPAAWATNGISTVWFWPSMARSSGSGVPSIRTAWSSKCKWCKAAEMPR
 RQSV"
 complement(1766..2200)
 /gene="ml16354"
 complement(1766..2200)
 /gene="ml16354"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB52666.1"
 /db_xref="GI:14026068"
 /translation="MVARHATIAEAVERTTPPSRVRRQAEHAPVSAKEALQALPED
 AWHMVTREGSKASLAFATVVRVPAHRDYWRSIPHAEEWFLIEWDPGEVEPTKYWL
 STLPKTRLADLVQAKMWRTERDYQELKQGCCHIDFGLTV"
 complement(1939..2931)
 /gene="ml16355"
 complement(1939..2931)
 /gene="ml16355"
 /codon_start=1
 /transl_table=11
 /product="transposase"
 /protein_id="BAB52667.1"
 /db_xref="GI:14026069"
 /translation="MLHMEGAMDLDGIETSESRTAAVETILASVLGHADRAVPLKAY
 CTGLILIPREKSVPEPAAVEPRGVAHQSLHHEFADKADSDDAVLSRVREQLPAL
 ERGOIIRAWIVDDTGFPRKKGSHVGVQYQGLQKDNQCVAVSLSVATQASLPPIA
 YOLIPESWANDPDRRAKTVGVNVVFTKPEIALQIRAALEAGVSRGVVLADAGYG
 NDTAFRTGLTKVGLTVYGVQSSIRLWSPGTQPLPKPWSGRCGPRPRAFGRLNMRRS
 QRANWKRCPRTPTGTLWLGARAPRRRLPHALPPCGGRIATTGAPSPTRLNGS"
 complement(3398..3616)
 /gene="msl6356"
 complement(3398..3616)
 /gene="msl6356"
 /note="unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BAB52668.1"
 /db_xref="GI:14026070"
 /translation="MLSGMPPCRLTMMAGATIDLLAQOAFVAGFOIETRTARHH
 RSAAKRDRSPGLRLNTNSARPADRSL"
 4196..6634
 /gene="mlr6358"
 4196..6634


```
/gene="mlr6358"
/notes="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB52669.1"
/db_xref="GI:14026071"
/translation="MADTSAPIHQRQVLVYIGDEKMTGIGRSGKSRIGDAGAGSSSSR
SQSLGASGEGSGQFGLVLRMRSEGGKSPSAGAARGAHDLDGKSSGSPVPPHPIPA
AAPRAAGPRAAGSRSSVSSTSDGALAPVASOPSTSGVTVPSSSGEVINETKRM
CSLIDELQVLEEKSLHSGKPMEGFDELQDOLANAGAELECYASLPKDRARSTLSDDD
VRYRADVTHRADSANELAHITFKKEQNRKRMEDILHNLVKSNTATLDLQSGVNSKIM
KYVACVERWKKVQKHELMKAVCAVADPLDPSGTAEMREEREAALRLHSGVNSKIM
QLOSRVQNNAPQAIMLEAPIRELLIDGQNTRELLGNFICEVPAFVTAELVR
KEGRDLDAECALVEGWELRADFASASHVVDRLDHQAGAGLPLELLTRIVDDAWI
ASDEVTDLLAQPKPAPDLTADAAMAPTGTAGKTMVAESTTAPRKKDKSKOAAAG
AGSSAAGRPEPOVAAVSDPAPAAKVLVRLDTGLTKLVSAEEAONASASAAARLALO
NPTTGENTKSLVARSTELLKFDLPQAQSLISRSRSGKPEDVEHAGLVVQRLQOAVE
MEPCLPALEASRLSKGKDDVHRLTVQFRLMLANVRGLAKAWDKQKPMIDCMKRY
PFSQNHLEQSLARELTPADPCALNPEPSNLFELRLQPKALSNTLPSPMVHIHT
TRKIYAOQLATDLSAEFAACHVKSNEQRHNRQWQTSRAKGYDNAIHRKGLTPAFCK
SLSLSDRMRNGPA"
complement(6985..7716)
/gene="ml16359"
complement(6985..7716)
/gene="ml16359"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB52670.1"
/db_xref="GI:14026072"
/translation="MRTWAMSEVARDPLYRCHRFPAEIIAHAVWLYFRPPLSRMVED
MLAAGIIVHTQITRISWAEKFGHREIKRSAGCLGDKWHLDECYVAINGKKQWLW
RAVDQGLVLEVLVQSRNAKAKCLMRKLLKAAQGTPRVMTDKLRSVDAARRDLP
GVHRRSHKGLNNRNSHOPTRRRERTMRKRSARQLQREVSIIHDP IANLHFFPNTL
SSAQRDLRANAMQIWNKIACLA"
8086..17256
/gene="mlr6361"
8086..17256
/gene="mlr6361"
/notes="similar to shikimate kinase"
/codon_start=1
/transl_table=11
/protein_id="BAB52671.1"
/db_xref="GI:14026073"
/translation="MTGVNRSRSTREKRVGRTGEASSRSOSSILAPGAGSOFDTVV
EQMSVEIINRTPAVPMERGAADLHDPGSPVSGVTHGTLDAAHAAAPRAAGPRR
SVPMLEGETLAPVASOPSTSGVAAPICLSGDEIKERKMDSLTLVGMNRNAARODP
GDEKLLDOVINAGSAOPHTKSLSPALARTIVDDQLITCRYLWDVAQVQESRLAEP
IHKLOKKAEBEALGLTSTATPESVVENHACVWRKKGCHGERGRAARATASLP
SATAKMRQEEATRLRATGVSVMHSLVLYLOSRLIDARAKIEPQASTFDPAVREILME
NGRVLLEAFSGDVLPAFLSTQNAVLSGGHPLDADHCAVLEGVIERLSEFASQLVI
RTNLNDHOTSADLPLELLSOIVEGAWIVGHEVHMLLELPKPPAVLPDAGAAIPAE
TAGSKSVAKGTALQPKAPALIASPADAGATIQIDTVGKSPVAKGTLALOPNWPANIA
PADMGAGIPADTAGKAAGAAARRRKGGKPAVGAGSSAPGRASQVAVSDSTAP
ADKVLVLDPMGKTKLSVAEARSASAGASQAKOAPPSEALTOERQVQKPEDARYAET
VERLOQTQEMRACPAALDEPRPGLLAPQAVEPQDKTVRLKMLNQAQELAKALN
SSSHAPGSSQSDAGVERTHSGPDRMSYARVVRGDEALGDKSADYSIAPRGALA
SGRGATSGPRHSGPQGRHEMGROESSILRLERSWLGQEPSRSDPFLARHDOSOR
PVAGPERRRFHVFPWEDDLDLNLHKIFPAHDINEESVAVVENDKRLQSOAHRT
FLRKAASEPTREHFVPRWEDDWTGATSCNATSEPGQAGMEACRAMAAQVSR
LGRALNDVESKTLISLVLSFSRHPRLACRNGMITIAKLEHQEPGLMEASQLALL
VNGFSKPEQENARKATVAAAEIRRRPAGLSDEPDQNLNVNGLSKWLEEECCGV
IDAIVKCGVRKAGTGLSGNSQAFANLVNFGSKAREPNSHQATIVADAVRLAD
ETVIASTIGRLADEKAGRTGWYPQALANLVNFGSKWPLDCRGAIVAIARVAGAGKL
PVFDAQALANLVNFGSKWPEDCRGATIAIAAEVCLADLPKAGLSGDPQHLANLVN
FSLKQAAECRSATVAIAGEVCHRRKGLSDFOVHLANLVNFGSKWQPVNLYGAI
NVVAGEVLRCAEPLSEFLPLHLASLVNFGSKWPRWNSHQATVAGVRHNRQDLSG
FEPQELANLVNFGSKWPEDCRGATVAVAPRRRAGRUSGFGQDLANLVNFGSKWPL
DCKEAWAIVARVAPRRRGLSGFNGQDLANLVNFGSKWPEQNSROATVATAGVLR
TNPPDFPSQVLEAVNFGSKWPEEELRQAAAATAGEVPIRSANRLPDTFSOHLNV
NGFSKWQOEKSRQATAAIAAGEVLAERLGSVNPQOALNLLNGFDRFABEACSOAI
LEIARLGOAGOPPRHFAAPGLSSTANSILARGILRSEDAIEAELKDLRLKLAHY
LHYASDRLEADAVGVTSLKALAKAQLYDDLGLSAGLNLRLAELHKAFAHNNL
```

```
alignment_scores:
  Quality: 112.00      Length: 264
  Ratio: 0.896        Gaps: 14
  Percent Similarity: 47.348  Percent Identity: 22.348

alignment_block:
US-09-674-779-2 x AP003009 ..
Align seg 1/1 to: AP003009 from: 1 to: 339681

5 AsnGlnTyrPheIleThrThrLeuLeuSerSerMetLeuValAlaCys.. 20
||||| : : : : : : : : : : : : : : : : : : : : : : : :
134211 AACCAAGATCGTGTTCATCAGCGGTGCGAGGAGCGATGCCCGCTGCCT 134260
21 ....SerAlaProIleProThrAsnProGlnValSerProIleLysThrP 36
: : : : : : : : : : : : : : : : : : : : : : : :
134261 GGTGAACCTCGCTGTCAGGTTTCGAAAGAGGCTCGCGCTT.....C 134304
36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGlu 52
||||| : : : : : : : : : : : : : : : : : : : : : : : :
134305 CGCGGAGCGTCAATCGTGGCAAGACAAATGTCAAGGCGATGAC 134354
53 HisAsp.GluSerValSerHisValGlyLeuGlnAlaHisPheGluThrT 69
: : : : : : : : : : : : : : : : : : : : : : : :
134355 CGTGATTGAAGCGAAGACAGACGCGCGCGCGCTCCCATCGTTGGACAT 134404
69 rpleu.....
|||||
134405 GGTTCGGCGCATGTCGTCGCTTCTTTGCTGTGCGTCACCGCGCG 134454
71 .....GlnMetHisAlaThrly 77
134455 TTCGCCGAGCGCGTCCCTTGAAGATCCAGCATCTGCATACGGCGGAGAA 134504
77 sGlnGluValVal.....ArgTyrGlnAlaTyrLeuGlnSerArgLeuG 92
| : : : : : : : : : : : : : : : : : : : : : : :
134505 GCGCGAGATCGTCTTCAAGCGCAACGACGCTATGATCAGCGGGATTG. 134553
92 lyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrp 108
: : : : : : : : : : : : : : : : : : : : : : : :
134554 .....AAGAAGATCGATTTCATGCTGCGCGACTGG 134583
109 GlnAlaCysGlyHisGlnProTyrGlnLeuProGluHisLeuTrpGl 125
: : : : : : : : : : : : : : : : : : : : : : : :
134584 CGCGCG.....AACGCGCGACGAGGATGGATCCGCGCTGCTC..... 134622
125 yGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyI 142
: : : : : : : : : : : : : : : : : : : : : : : :
134623 .GATCTCGTC.....TGGCAGGCTATCGCGCGCGCGG. 134655
142 leuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu 158
|| : : : : : : : : : : : : : : : : : : : : : : :
134656 ..TCGAGCGCTATATCCATGTCGTCAGCGCTATCGCTCGCGCGGACC 134703
159 Asn.....GlnCysAlaGlyGlyAlaAlaMetSerLysHi 170
|| : : : : : : : : : : : : : : : : : : : : : : :
134704 AAGCCATCTCGCGCGCGCTCGAAGGCGTGGCGGGAAGCCAGCA 134753
170 sLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluLeuLys 187
: : : : : : : : : : : : : : : : : : : : : : : :
134754 CATGTGCGCGCGCGCTATCTTCTCGCGGACGCTCGCGCTGAAGA 134803
187 erGlnAlaLeuTyrGlnLeuGlnAsnArgLeuCysGlnTyrTrpLeuGlu 203
: : : : : : : : : : : : : : : : : : : : : : : :
134804 AG.....CTGCGCGACATCGCGCTCAAGATG..... 134829
204 HisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAla.. 219
|| : : : : : : : : : : : : : : : : : : : : : : :
134830 .....CAGGGCGGTGCTTGTGTTATTATCCGACCTCCGCGCTC 134867
220 .....IleHisLeuAspThrGlnGlyPheArgLysTrp 230
```

```

134868 GCCTTTCATTCACATGACGTCGGCATGTCGGCATGG 134907
seq_name: gb_htg:AC046147

seq_documentation_block:
LOCUS AC046147 120405 bp DNA linear HTG 16-OCT-2001
DEFINITION Mus musculus chromosome 12 clone RP23-265L9, *** SEQUENCING IN
PROGRESS ***, 38 unordered pieces.
ACCESSION AC046147.5 GI:16118086
VERSION AC046147.5
KEYWORDS HTG; HTGS; PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 120405)
Metzker,M.I., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
Gorell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hognes,M.,
Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
Perez,L., Reiter,D., Sav,J., Shen,H., Vasquez,L., Watlington,S.,
Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstein,G.,
Worley,K. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 120405)
Worley,K.C.
Direct Submission
Submitted (13-APR-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:11094635.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: MADW
Center clone name: RP23-265L9
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 147066 bases at least Q40
Consensus quality: 188973 bases at least Q30
Consensus quality: 206320 bases at least Q20
Estimated insert size: 186499; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 5655: contig of 5655 bp in length
* 5656 5755: gap of unknown length
* 5756 10324: contig of 4569 bp in length

```

FEATURES

Location/Qualifiers

```
source      1..120405
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="12"
            /clone="RP23-265L9"
BASE COUNT  35100 a 22309 c 23576 g 35674 t 3746 others
ORIGIN

alignment_scores:
  Quality: 110.00      Length: 208
  Ratio: 1.310        Gaps: 10
  Percent Similarity: 40.385  Percent Identity: 23.558

alignment_block:
US-09-674-779-2 x AC046147/rev ..

Align seg 1/1 to reverse of: AC046147 from: 1 to: 120405

48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
91957 CACCACACACACACACACACACACACACACACACACACACACACAC 91908

64 aHisPheGluThrTrpLeuGlnMethHisAlaThrLysGlnGluValV 81
:|||||
91907 ACAC.....ACACACACATA..... 91892

81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
|||||
91891 .....TATCTGCC..... 91883

98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisG 114
91883 ..... 91883

114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrL 131
91883 ..... 91883

131 euHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThr 147
|||||
91882 .....CAGGATTTCAGTAGAGACCCATGCTACCAACCCAGC... 91847

148 GlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyG 164
:|||||
91846 ...CTGAGTTGACGTGGTCTGGGACCAACATCTCTATGC..... 91808

164 yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp...IleTrpV 180
|||||
91807 ....TTGCCTGACAAAGCACCCTTACCACCTGAGCTATCTCCCAAGCTGGG 91762

180 alProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeu.....Gln 194
|||||
91761 TACCACACACTAAATTAGAGTTTGTAGACTTAGTAAAGCTTATGACAAGAG 91712

195 AsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhel 211
:|||||
91711 GACAGGAGTGGGAGTGTGGCTGGAACACAGCAAAATCTGAATGTGGC 91662

211 yLeuGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe. 227
|||||
91661 TTCAGT.....TGACAACTACATTTCACACCTTTT 91630

228 .....ArgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCys 241
:|||||
91629 CAAGCCGGGCATCATGG.....GTATTCTCAGATCTATGGGCAATGT 91586

242 ArgHisValLeuProLysAsnLys 249
|||||
91585 ACTCACATATCTCTGAAACAAAG 91562

seq_name: gb_htg.AC073821
```

seq_documentation_block:

LOCUS AC073821 145347 bp DNA linear HTG 02-SEP-2000
DEFINITION Mus musculus clone RP23-79P7, WORKING DRAFT SEQUENCE, 10 ordered
pieces.
ACCESSION AC073821 GI:9964818
VERSION AC073821.2 GI:9964818
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 145347)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
2 (bases 1 to 145347)
DOE Joint Genome Institute.
DIRECT SUBMISSION
SUBMITTED (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 2, 2000 this sequence version replaced gi:8810438.

COMMENT

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 1767575
Center clone name: RPCI-23-79P7

Summary Statistics

Consensus quality: 138992 bases at least Q40
Consensus quality: 143251 bases at least Q30
Consensus quality: 144092 bases at least Q20
Estimated insert size: 150000; agarose-fp estimation
Estimated coverage: 9.13 in Q20 bases; agarose-fp estimation
Quality coverage: 9.45 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 20853: contig of 20853 bp in length
20854 20953: gap of unknown length
20954 54936: contig of 33983 bp in length
54937 55036: gap of unknown length
55037 75800: contig of 20764 bp in length
75801 75900: gap of unknown length
75901 79363: contig of 3463 bp in length
79364 79464: gap of unknown length
79464 82638: contig of 3174 bp in length
82638 82737: gap of unknown length
82737 107168: contig of 24431 bp in length
107168 107268: gap of unknown length
107269 109162: contig of 1894 bp in length
109163 109263: gap of unknown length
109263 136238: contig of 26976 bp in length
136239 136239: gap of unknown length
136239 137999: contig of 1661 bp in length
138000 138099: gap of unknown length
138100 145347: contig of 7248 bp in length.

FEATURES

source

1..145347
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-79P7"

BASE COUNT

37072 a 35595 c 34914 g 36864 t 902 others

Tue Sep 17 07:27:30 2002

ORIGIN

```

seq_name: gb_htg:AC073767
seq_documentation_block:
LOCUS AC073767 201377 bp DNA linear HTG 18-JUL-2000
DEFINITION Mus musculus clone RP23-359H6, WORKING DRAFT SEQUENCE, 13 ordered
pieces.
ACCESSION AC073767
VERSION AC073767.2 GI:9256787
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 201377)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201377)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:8810384.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1874902
Center clone name: RPCI-23_359H6
-----
Summary Statistics
Consensus quality: 189664 bases at least Q40
Consensus quality: 197375 bases at least Q30
Consensus quality: 198999 bases at least Q20
Estimated insert size: 176970; agarose-fp estimation
Estimated insert size: 200827; sum-of-contigs estimation
Quality coverage: 8.42 in Q20 bases; agarose-fp estimation
Quality coverage: 7.42 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 9998: contig of 9998 bp in length
* 9999 10098: gap of unknown length
* 10099 17043: contig of 6945 bp in length
* 17044 17143: gap of unknown length
* 17144 40422: contig of 23279 bp in length
* 40423 40522: gap of unknown length
* 40523 70835: contig of 30313 bp in length
* 70836 70935: gap of unknown length
* 70936 82602: contig of 11667 bp in length
* 82603 82702: gap of unknown length
* 82703 97050: contig of 14348 bp in length
* 97051 97151: contig of 19755 bp in length
* 97151 116905: contig of 17005 bp in length
* 116906 132589: gap of unknown length
* 132590 132690: contig of 15584 bp in length
* 132690 135229: gap of unknown length
* 135230 147912: contig of 2540 bp in length
* 147913 148012: gap of unknown length
* 148013 159289: contig of 12583 bp in length
* 159290 175515: gap of unknown length
* 175516 175515: contig of 11277 bp in length
* 175515: gap of unknown length
* 175516 175515: contig of 16126 bp in length
* 175516 175515: gap of unknown length

```



```

/gene="mll12564"
complement(6734. .7366)
/gene="mll12564"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49663.1"
/db_xref="GI:14023056"
/translation="MPSTELLIAFAFATAIFAYIPGPAMLYAAAOTLARGRWSGLTAA
LGIHLGGYVHFAAAGLSLHFAVPPHLMKVLIGALYLWLGVSIFPKRGVEGDVAL
PAIERKSARRAFFSITIVELNPKTAIFFMAFLPQFIDASAALPFWLQFVVLGTIYNL
PFESADIYCVFLAGLIVTSRLRRSRAQRVMQRAGGAVLGLGVHVAFLQKS"
7499. .8449
/gene="mll12565"
7499. .8449
/gene="mll12565"
/codon_start=1
/transl_table=11
/product="ADP-L-glycero-D-mannoheptose-6-epimerase"
/protein_id="BAB49664.1"
/db_xref="GI:14023057"
/translation="MLIVTGAGMIGSNIVAALNAEGHDDILVVDLTDGCHKIANLAD
LSIADYLDKKDPLRIEACALGRTEAVPHQACSTTEWNGKEMVEYNFAYSKRLLHA
COAPKVPFLYASSASVYGGGSEFEPEPALERPLNVIYAYSKLFDYVRRVTFDTHSQ
VAGLIRFVNYGPREAHKAMASVAFHLFNVQVERGENPOTFLGAYDGFQGGQSRDFHIV
GDVADYNLWMKRGSGSFGNCTGRAQPFRAIAETVIDTLGKGIEIFIPFPDHUKGSI
QSFTQADMSRLRAAGYNGQFRTVETGRVDYVEWLKAQRSS"
8422. .9429
/gene="mll12566"
8422. .9429
/gene="mll12566"
/codon_start=1
/transl_table=11
/product="Adp-heptose-LPS heptosyltransferase"
/protein_id="BAB49665.1"
/db_xref="GI:14023058"
/translation="MAESTILVIGPRVGDWVMAQCLFSALKELYPNAAIDVLAPAW
APFLVRMPEIRQQLDPLPKGALEFTRRGRGLRGYDMAYVLPGSKWSALIPPE
ARLPDRFGLRMRYGLDLDIVLPDPAVKRTAQAYFSLAKGGSFRAFLHIVDAGNO
ARTPRFGLAPRKFAAPGAEPKARWPFSEYAGLAKFKMDKGLKVALFGSKNDRO
TABITAPGVVDLAGQTRLEDAIDIIAARLAVSNDSLMHVAAGVGTPIVAVYGS
SPENTPLAERRELVWLHSCSPCHQKVCPLGHLNCLKTLEVGQVATAAERLLELPAT
A"
9426. .10400
/gene="mll12567"
9426. .10400
/gene="mll12567"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB49666.1"
/db_xref="GI:14023059"
/translation="MKVLIVKTSMSGDVTHTFPAVEDASLHRPDISFQWVEEAFAGI
VALHPATAGHTHTVAIRWRKALFNNGTWEAAALRTRURECFRDLIVDAQGLLKSA
ARQAPAIAGFDRSSAREPSATILFYDTYGVCPNRLHAERTRELFLGALGYOPDLST
ASGVTPPGALGICTKTAFLHGTSTRDKWPVEDWIGTARLLVERGMTPTVYTMSN
REKSVAEAIRAVPSTVVVYKSPADIAILIGRSTLVIIGADTGLTHLASAFGLPTVA
FLATEPGLTGPGRPYASTLLAVPGRLTAEVVVAEADSLNLKIQAPA"
complement(10535. .11008)
/gene="mll12568"
complement(10535. .11008)
/gene="mll12568"
/codon_start=1
/transl_table=11
/product="transcriptional elongation factor"

```

Align seg 1/1 to reverse of: AP003000 from: 1 to: 348077

```

72 MetHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuG1 88
339623 CTGCACACGCATGAAGAGCCGAGATCGTC.....TA 339592

88 nSerArgLeuGlyAsnTyrLeuProPro...MetSerGlnLeuLeuThrt 104
339591 CAAGCGCAACGCCGCTACCTCCCGAGGCTCGCAAGATCAACATCA 339542

104 hrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPro 120
339541 TCTCGCGCAGCTGGCGCGC.....AAGGAGCCGACCAAGATGATCGG 339498

121 GluHisLeu.....TrpGlyGlnIleValProThrLeuHisLeuTyr 134
339497 CGCGTCTGGAGCTGCTGCG.....GAGGCGCTA 339469

134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 151
339468 TCGGAAAGCGCGCCGACGACTATATC.....CAGGTGCGGT 339431

151 erValTyrArgAsnProGluLeuAsn.....GlnCysAla 162
339430 CGCGTACCTGCTCGCGCGGACCACTCGATGCTCGCGAGCGCGCGC 339381

163 GlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTr 179
339380 GGTGTCGCGGAGAGAGCAGCATATGCTCGGAGCGGATGGACTCTA 339331

179 pValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsn 196
339330 CATTCGCGCGTGGCGCTGGAAGAGCTCGCAATATCGCGCTG..... 339288

196 rgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeu 212
339287 .....AAGATCGAGCGGCGGTGGCGTC 339267

213 GlyLeuTyrAlaThrGlyAla.....IleHisLeuAspThrGlnG1 226
339266 GGCTACTATCGAGCTCGCGGTTCGCCATTCGTCATATGATGTCGCGCA 339217

226 yPheArgLysTrp 230
339216 TGTGCGCCACATGG 339204

seq_name: gb_pr:AC068614

seq_documentation_block:
LOCUS AC068614 164706 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-512N1 from 2, complete sequence.
ACCESSION AC068614
VERSION AC068614.5 GI:15145605
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 164706)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE
2 (bases 1 to 164706)
AUTHORS Kang,K., Abbott,A., Boyer,E. and Dixon,R.
TITLE The sequence of Homo sapiens BAC clone RP11-512N1
JOURNAL Unpublished (2001)
REFERENCE
3 (bases 1 to 164706)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

REFERENCE

AUTHORS
TITLE
JOURNAL

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

MO 63108, USA
4 (bases 1 to 164706)
Waterston,R.H.
Direct Submission
Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 164706)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced gi:14626397.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_NH0512N01

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Cataneese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-102G8, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-512N1; actual end is at base position 164706 of RP11-512N1.

Polymorphisms exist between RP11-512N1 and RP11-102G8. Data from AC032014 was used to finish this clone AC068614.

FEATURES

source

1. 164706
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"

repeat_region

1. 138
/clone="RP11-512N1"

repeat_region

145_804
/rpt_family="MER1_type"

repeat_region

836_1669
/rpt_family="L1"

repeat_region

1685_1848

Tue Sep 17 07:27:30 2002

```
repeat_region /rpt_family="MER1_type"
1868. .2044
repeat_region /rpt_family="MER1_type"
2260. .2411
repeat_region /rpt_family="CR1"
2485. .3182
repeat_region /rpt_family="L2"
3172. .3293
repeat_region /rpt_family="L2"
3298. .3488
repeat_region /rpt_family="L2"
3458. .3484
repeat_region /rpt_family="AT_rich"
3617. .3718
repeat_region /rpt_family="MER1_type"
3719. .3886
repeat_region /rpt_family="Alu"
3887. .3946
repeat_region /rpt_family="MER1_type"
3954. .4018
repeat_region /rpt_family="MIR"
5044. .5194
repeat_region /rpt_family="MIR"
5700. .5780
repeat_region /rpt_family="MER1_type"
6317. .6551
repeat_region /rpt_family="MIR"
6824. .6871
repeat_region /rpt_family="CT_rich"
7157. .7256
repeat_region /rpt_family="L2"
8251. .8715
repeat_region /rpt_family="MaLR"
8744. .8909
repeat_region /rpt_family="L2"
8933. .8971
repeat_region /rpt_family="MIR"
9722. .9786
repeat_region /rpt_family="L2"
10431. .10489
repeat_region /rpt_family="L2"
10886. .10897
repeat_region /rpt_family="MIR"
12174. .12338
repeat_region /rpt_family="L1"
12430. .12505
repeat_region /rpt_family="L2"
12688. .12963
repeat_region /rpt_family="Alu"
13850. .13945
repeat_region /rpt_family="MIR"
14692. .15044
repeat_region /rpt_family="L2"
15938. .16906
repeat_region /rpt_family="L1"
17595. .18092
repeat_region /rpt_family="Achobo"
18081. .18507
repeat_region /rpt_family="Achobo"
18374. .18426
repeat_region /rpt_family="AT_rich"
19734. .19816
repeat_region /rpt_family="MIR"
19971. .20124
repeat_region /rpt_family="MER103"
20810. .20877
repeat_region /rpt_family="(TGGA)n"
21098. .21164
repeat_region /rpt_family="L2"
21865. .22233
repeat_region /rpt_family="L2"
22328. .22439
repeat_region /rpt_family="MIR"
22440. .22664
repeat_region /rpt_family="MIR"
22665. .22747
repeat_region /rpt_family="MIR"
23939. .24091
repeat_region /rpt_family="MIR"
24668. .24701
repeat_region /rpt_family="(TTTA)n"
24783. .25084
repeat_region /rpt_family="Alu"
25063. .25084
repeat_region /rpt_family="AT_rich"
25255. .25297
repeat_region /rpt_family="L2"
26227. .27186
repeat_region /rpt_family="ERV1"
27187. .27284
repeat_region /rpt_family="ERV1"
27285. .27610
repeat_region /rpt_family="ERV1"
27848. .28178
repeat_region /rpt_family="L2"
29215. .29249
repeat_region /rpt_family="AT_rich"
30350. .30386
repeat_region /rpt_family="AT_rich"
32688. .32732
repeat_region /rpt_family="MIR"
33373. .33665
repeat_region /rpt_family="(TA)n"
35134. .35187

alignment_scores:
Quality: 106.00 Length: 260
Ratio: 0.855 Gaps: 12
Percent Similarity: 47.692 Percent Identity: 22.308

alignment_block:
US-09-674-779-2 x AC068614/rev ..
Align seg 1/1 to reverse of: AC068614 from: 1 to: 164706

8 phe1leThrLeuSerMetLeuValAla.....CysSerAl 22
|||||: ||| |||||: |||||: |||||:
95720 TTTTCTCTCTCTGGCTCCCAACCTCTGATCTCTCCCATTCGCTATA 95671

22 apToleProThrAsnProGlnValSerProLleLysThrProSerValI 39
|||||: |||||: |||||: |||||:
95670 CCCACTCCAGTCACCCCC.....TTTA 95648

39 eu1leThrLysAspLysIleGlyAspHisThrHisGluHisAspGlu 55
|||||: |||||: |||||: |||||:
95647 TATACACATACACAAATTGAGGACAGTCATCTTGTAGTCAGACAAA 95598

56 SerValSer.....HisValGlyLeuGlnAlaHisPheGluTh 68
|||||: |||||: |||||: |||||:
95597 GAAGAAATTTTCGCTGAAGGCACACAGCTCAAGAGGTCAT..... 95556

68 rTrpLeuGlnMethHisAlaThrLysGlnGluValValArgTyrGlnA 85
|||||: |||||: |||||: |||||:
95555 TGTCATCCAGATCCCAATGAGAGAGGAGTTTCTGTCCCATATAAGA 95507

85 latYrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGlnLeu 101
|||||: |||||: |||||: |||||:
95506 GGTTCG.....CTGGAATGAGAGTTTCCCATCTTCAGGAGCT 95467

102 LeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLe 118
|||||: |||||: |||||: |||||:
95466 TTC.....TCTGGGAAGTTGCAGACACACACAGGATGAGTTG 95429

118 uproProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrG 135
|||||: |||||: |||||: |||||:
95428 GTCACCCGTCATCTAGGCAAGCTGGTGAACAGCTTTTATTCTGTGTTT 95379
```



```

135 lnAspLeuIysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
    :||| ||||| ||||| :||| :|||
95378 ATCTTGCTATCCAGCATTTATCTTGAGCCTCAGGTTCA...AGGCAT 95332
    :||| :||| :||| :||| :|||
152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
    :||| :||| :||| :||| :|||
95331 ACATTTAGAGAGAGAGAGTGTACAGGCTCTCTTGTGTGTCTGTCTTACA 95282
    :||| :||| :||| :||| :|||
168 rLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluI 185
    :||| :||| :||| :||| :|||
95281 AATCAAGAAGAAGAACAGATATTACAGTG...ATTTCACACCTCCCAA 95235
    :||| :||| :||| :||| :|||
185 lelysSerGlnAlaLeuTyrGluLeuGlnAsn.....ArgLeuCys 198
    :||| :||| :||| :||| :|||
95234 TTACTTTTTCACACCTCCCTCCACATCCAGCATCCACATGGAAGTTGTGT 95185
    :||| :||| :||| :||| :|||
199 GlnTyrTrpLeuGlu.....HisGlyGluAsnGlnAsnPheGlyLeuG 213
    :||| :||| :||| :||| :|||
95184 AATGGTATGTGGAATCACATCCTCC..... 95158
    :||| :||| :||| :||| :|||
213 yLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLys 230
    :||| :||| :||| :||| :|||
95157 ....TATGCTCAGAGTCAGGTTTCAGTGGACACACATGTGGTGCTCCT 95112
    :||| :||| :||| :||| :|||
230 rp.....GlyAlaGlnPheSer 235
    :||| :||| :||| :||| :|||
95111 GGGAAGATAAAGAAATAATTTCTGTTGGAGGAGGAGGTAGCAAAACA 95062
    :||| :||| :||| :||| :|||
236 GluThrAsnSerIleCysArgHisValLeu 245
    :||| :||| :||| :||| :|||
95061 GAAATAATTTCTTTGTCAAGCACATAATG 95032
    :||| :||| :||| :||| :|||
seq_name: gb_htg:AC098619
seq_documentation_block:
LOCUS
DEFINITION
Rattus norvegicus clone CH230-125E15, *** SEQUENCING IN PROGRESS
***, 82 unordered pieces.
ACCESSION
AC098619
VERSION
AC098619.3 GI:17973396
KEYWORDS
HTG; HTGS-PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 184172)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bivaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Duan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hughes,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.B., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

```

```

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 184172)
Worley,K.C.
Direct Submission
Submitted (27-Oct-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064347.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHVG
Center clone name: CH230-125E15
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 147975 bases at least Q40
Consensus quality: 158603 bases at least Q30
Consensus quality: 166972 bases at least Q20
Estimated insert size: 151344; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 82 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
*
* 1 5267: contig of 5267 bp in length
* 5268 5367: gap of unknown length
* 5368 9333: contig of 3966 bp in length
* 9334 9434: gap of unknown length
* 9434 14354: contig of 4921 bp in length
* 14355 14455: gap of unknown length
* 14455 18618: contig of 4164 bp in length
* 18618 23127: gap of unknown length
* 23127 23227: contig of 4409 bp in length
* 23227 25777: gap of unknown length
* 25777 25877: contig of 2550 bp in length
* 25877 28855: gap of unknown length
* 28855 32638: contig of 2978 bp in length
* 32638 32738: gap of unknown length
* 32738 37330: contig of 3683 bp in length
* 37330 37331: gap of unknown length
* 37331 40552: contig of 4592 bp in length
* 40552 40652: contig of 3122 bp in length
* 40652 43953: gap of unknown length
* 43953 44058: contig of 3306 bp in length
* 44058 44059: gap of unknown length

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```
* 131125 131224: gap of unknown length
* 131225 133021: contig of 1797 bp in length
* 133022 133121: gap of unknown length
* 133122 134279: contig of 1158 bp in length
* 134280 134379: gap of unknown length
* 134380 136332: contig of 1953 bp in length
* 136333 136432: gap of unknown length
* 136433 137878: contig of 1446 bp in length
* 137879 140375: gap of unknown length
* 140376 140475: gap of unknown length
* 140476 141842: contig of 1367 bp in length
* 141843 141942: gap of unknown length
* 141943 143652: contig of 1710 bp in length
* 143653 143752: gap of unknown length
* 143753 145714: contig of 1962 bp in length
* 145715 145814: gap of unknown length

alignment_scores:
    Quality: 105.50      Length: 241
    Ratio: 0.977        Gaps: 13
    Percent Similarity: 44.813    Percent Identity: 21.162

alignment_block:
US-09-674-779-2 x AC098619 ..
Align seg 1/1 to: AC098619 from: 1 to: 184172

5 AsnGlnTyrPheIleThrThrLeuIle..... 13
|||||:|||||:|||||:|||||:
36176 ATAAACATTTCTCTCCAGTTTGTAGTCTGGGGAGTTTATAGACTA 36225

14 ....SerSerMetLeuValAlaCysSer.....Alap 23
|||||:|||||:|||||:|||||:
36226 CTAAACGGGCATTATGTTGTATAGTCTTTGTCACCCAGCCTTAAGCTC 36275

23 rolleProThrAsnProGlnValSerProIleLysThrProSerValLeu 39
|||||:|||||:|||||:|||||:
36276 CAGTCACTACCAACCCCATCCAGATCCCTTGAATCCACAAGT..... 36319

40 IleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspLys 56
|||||:|||||:|||||:|||||:
36320 .....AAAAATCACACACACACACACACACACACACACACAC 36351

56 rValSerHisValGlyLeuGlnAlaHis..... 65
|||||:|||||:|||||:|||||:
36352 ACACACACACACACACACACACACACACACACTTATTTAAATGCCCTT 36401

66 .....PheGluThrTrp.....LeuGlnMethHisHisAlaThr 76
|||||:|||||:|||||:|||||:
36402 ATTAACCTCAATGACTGGGCACCTGTTAAGCTCTCTCCACACAGGCAAT 36451

77 LysGlnGluVal...ValArgTyrGlnAlaTyrLeuGlnSerArgLeuGl 92
|||||:|||||:|||||:|||||:
36452 TCGGATAGATGGCCATTGGCTACTGAATCAAAACTCACATGGCTGGT 36501

92 YAsnTyrLeu.....ProP 97
|||||:|||||:|||||:|||||:
36502 TGGCTTCATCTCTGAGTACTACTCTTTCTCTCTCTCTGACCCCC 36551

97 roMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHis 113
|||||:|||||:|||||:|||||:
36552 CT.....CAATCTCGCTGTGT 36568

114 GluProTyrGlnLeuProGluHisLeuTrpGlyGlnIleValProth 130
|||||:|||||:|||||:|||||:
36569 CCTCCTTAAGTGTCCACCCCTGCCCTCTG...TCACTGGTGTCTTGAAT 36615

130 rLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsn 147
|||||:|||||:|||||:|||||:
36616 TTTATTATGATCAAGAA.....CCAATTGGGG 36644

147 hrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGly 163
```

```

36645 AACAGGCTCTTAGTGTCT..... 36661
164 GlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVa 180
36662 ....AGAACATCTCTAAACAT.....TATGATGTGTATTG 36693
180 lProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgL 197
36694 TCCAGCATCCCATTAATAAGCAA.....ATTGAACACTGT 36731
197 euCysGlnTyrTrpLeuGluHis 204
36732 ATTTTAGCTATTGGCTACACTAT 36754

seq_name: gb_htg:AC020643

seq_documentation_block:
LOCUS AC020643 162527 bp DNA linear HTG 30-AUG-2001
DEFINITION Homo sapiens chromosome 03 clone RP11-486C24, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC020643
VERSION AC020643.4 GI:9887661
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Smith,D.R.
Genome Therapeutics Corporation Sequencing Center: Human Genome
Sequence Data
Unpublished
2 (bases 1 to 162527)
Smith,D.R.
Direct Submission
Submitted (07-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
Street, Waltham, MA 02453, USA
On Aug 24, 2000 this sequence version replaced gi:8389446.
-----
Genome Center
Center: Genome Therapeutics Corporation
Center code: GNC
Web site: http://www.genomecorp.com/
Contact: gtc-seqcenter@genomecorp.com
-----
Project Information
Center project name: hg085
-----
Summary Statistics
Sequencing vector: N/A
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 990315
Consensus quality: 149308 bases at least Q40
Consensus quality: 156323 bases at least Q30
Consensus quality: 157579 bases at least Q20
Insert size: 161926; sum-of-contigs
Quality coverage: 5.6x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2231: contig of 2231 bp in length
2232 2231: gap of unknown length
2332 7785: contig of 5454 bp in length
7786 7885: gap of unknown length
7886 20723: contig of 12838 bp in length
20724 20823: gap of unknown length
20824 41799: contig of 20976 bp in length
41800 41899: gap of unknown length
41900 62173: contig of 20274 bp in length

```

```

* 62174 62273: gap of unknown length
* 62274 93394: contig of 31121 bp in length
* 93395 93494: gap of unknown length
* 93495 129282: contig of 35788 bp in length
* 129283 129382: gap of unknown length
* 129383 162527: contig of 33145 bp in length.
FEATURES
    source
        1..162527
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="03"
            /clone="Rp11-486C24"
            /clone_lib="RPC1-11"
            1..2231
                /note="assembly_name:Contig6"
            2332..7785
                /note="assembly_name:Contig7"
            7886..20723
                /note="assembly_name:Contig8"
            20824..41799
                /note="assembly_name:Contig9"
            41900..62173
                /note="assembly_name:Contig10"
            62274..93394
                /note="assembly_name:Contig11"
            93495..129282
                /note="assembly_name:Contig12"
            129383..162527
                /note="assembly_name:Contig13"
            clone_end="7"
BASE COUNT 49414 a 32608 c 33570 g 46221 t 714 others
ORIGIN

alignment_scores:
    Quality: 103.50      Length: 286
    Ratio: 0.755        Gaps: 14
    Percent Similarity: 47.902    Percent Identity: 18.881

alignment_block:
US-09-674-779-2 x AC020643/rev ..
Align seg 1/1 to reverse of: AC020643 from: 1 to: 162527
15 SerMetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSe 31
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74680 ACTCTCTGGCAAGCTGAAGCCACATTCCCTCCATCCAGCCTCCACG 74631
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
31 rProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspH 48
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74630 CCCCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 74600
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
48 iHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAla 64
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74599 ..CACACACACACACACACACACACACACACACACACACACATGCT 74552
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
65 .....HispheGluThrTrpLeu.....G1 71
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74551 TAGGAGAAACAGAACCCAGAGAGATTTCAGACATTTATTTGCTTTGGCA 74502
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
71 nMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuG 88
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74501 AATCTACCACTCACCTAAGAACCCAGGACTTTTGTGCT.....TTAG 74458
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
88 InSerArgLeuGly.....AsnTyrLeuProMetSer... 99
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74457 AATCTCAGATGGAAAGGACCCCTAAGTATTATTATTTGTTTTTCATGA 74408
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
100 .....GlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysG1 112
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::
74407 ATTCTTTTATTCTGTACTTAGTCAGTCAGGATGGCAAGCATGTGG 74358
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::

```


155191 AAAAATGAGATTGTAAACATCACCTGGCAGCATTTTCATCAGATGAAGA 155

```

/translation="MRRKLLQWGLGAGLPLAARADDDIIGAL
PPPTTVAATVASIDPPALKPAPVDPWVHLHNVTGKLEAVYWG

```



```

CDS      /gene="CCI521"
          complement(11296..12432)
          /gene="CCI521"
          /note="similar to GB:D30759 PID:551631; identified by

alignment_scores:
  Quality: 103.00      Length: 237
  Ratio: 0.963         Gaps: 12
  Percent Similarity: 45.148  Percent Identity: 23.629

alignment_block:
US-09-674-779-2 x AE005826 ..
Align seg 1/1 to: AE005826 from: 1 to: 12528

28 ProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 CCTGAGCCACCGCGTTCGCGCGCGACACCGCTGCGTGGCGCGGTCCG 230
42 .....LysAspLysIleGlyAspHisThrH 51
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 TTCGATCGATCCGCGCGCGCTGAAGCTGTGTGATCCGCGTGGGTTC 280
51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
|||:|||||:|||||:|||||:|||||:|||||:|||||:
281 ATCTCCACAAC..... 291
68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 .....GTCCATACCGCGGAGAGCTGGAAGCGGTCTACTGGGA 329
84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuPro....PrometSerg 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 GAAC.....GGCGACTATGTTCGGACGCCGTGAGTG 361
100 InLeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyr 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 CGCTGGACAAGGTGTGCGGACTATCGCAACGACGAGGTCCTCCATCC 411
117 GlnLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLe 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 GATCGGGCCCTATGACCTCTCGACCGACCATCGCG..... 447
133 uTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleA 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
448 .....CGAAGACCCAGACGAGGCGCGTTC.....CAGTGA 481
150 rgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGly..... 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 TCTCGGGCTACCGCTCTCCGGCGCAACCGCTGCTGCCAAGCGCAGC 531
165 .....AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIle 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 GCGGAGGTGCCAAGAGACCTGCACATGATGGCAAGGCCATGGACAT 581
178 eTrpValProAspLeuGluLeuLysSerGlnAlaLeuTyrGluLeuGlnA 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 CTTCCTGGAGGATGTCGAGCTCAAGCAGCTCCGCGCTGCGGCTTTGGAT. 530
195 snArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGly 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 .....CTGAGCGTTGGC.....GCG 645
212 LeuGlyLeuTyrAlaThr...GlyAlaIleHisLeuAspThrGlnGlyPh 227
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
646 GTCGGCTACTCCGACGACCACTTCGTCATGTCGATGTCGGACCGGT 695
227 eArgLysTrp.....GlyAlaGln.PheSerGluThrAsnSer 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
696 CCGCAAGTGGACGGGACCTGAGGCGCTACTCGCCTCTCAGCCGGAATTGC 745
240 IleCysArg 242
|||:|||||:

```

746 ATAACCAAG 754

seq_name: gb_htg:AC003656_4

seq_documentation_block:

```

WPCOMMENT
Sequence split into 8 fragments LOCUS AC003656 Accession AC003656
Fragment Name Begin End
AC003656_0 1 110000
AC003656_1 100001 210000
AC003656_2 200001 310000
AC003656_3 300001 410000
AC003656_4 400001 510000
AC003656_5 500001 610000
AC003656_6 600001 710000
AC003656_7 700001 753000
Continuation (5 of 8) of AC003656 from base 400001 (AC003656 Homo sapiens clone PI C1

```

alignment_scores:

```

Quality: 102.00      Length: 326
Ratio: 0.823         Gaps: 20
Percent Similarity: 38.037  Percent Identity: 23.926

```

alignment_block:

US-09-674-779-2 x AC003656_4/rev ..

Align seg 1/1 to reverse of: AC003656_4 from: 1 to: 110000

```

16 MetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSerPr 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88874 ATGTCGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 88825
32 oIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH 49
|||:|||||:|||||:|||||:|||||:|||||:|||||:
88824 GCTA.....CACT 88817
49 isThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88816 CCTGCTCTGATCAACAACACACATGCTCGCGGAGGGG.....CGG 88776
66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValAr 82
|||:|||||:|||||:|||||:|||||:|||||:|||||:
88775 ACACAATGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 88743
82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPrometS 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88742 .TACCAGGCGCTTG.....CCGAGACTGGGATCCTGGCTGCTCCACCCCA 88700
99 erGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu... 114
|||:|||||:|||||:|||||:|||||:|||||:|||||:
88699 AGCCCCACAGGTCGGTGGGCTCTCTGTGACACTATGTGGAGAGGGCG 88650
115 ProTyrGlnLeuPro.....ProGluHisLe 123
|||:|||||:|||||:|||||:|||||:|||||:|||||:
88649 CCAGGTGGCTACCCAGTCTAGGTGGGACACAGCACTGAGGAGAAGCC 88600
123 uTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88599 CTGGGA.....TGCCCACTGAGCCTCTGTGTACAGCCCTCAGAAACC 88556
140 rgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr..... 153
|||:|||||:|||||:|||||:|||||:|||||:|||||:
88555 AC.....TGCCAGCTAACACACAGCCACCGGTTTCTGGGCGGAGGCC 88512
154 .....ArgAsnProGluLeuAsnGln.. 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88511 AGCCCTGGGGCCAGTGGCCCTCTCTGTAGCCGGCACCCAGAGGCATGCAAAG 88462
160 ..... 160
88461 AGGCCCCAGGGCTGGGCCCCAGGAGAAACACTGACCCCTCAGACCCACAG 88412

```


161CysAlaGlyValAlaMetSerLysHisLeuThrAsnSer	174
88411	GAGCCCTTTGTGGGAGGAGCTGCCTTGACGTCCTCGAACCA	88362
175	AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeu	190
88361	GGTGTCAACCCTCTGG.....GGCGTCAGACACAGGACGCC	88324
191TyrGluLeuGlnAsnArgLeuCysGln.	199
88323	AGGTGTGCTAAATGTCTCCCTGGCCCCCACCAGCATCTCCCTGTGGTC	88274
200TyrTrpLeuGlu	203
88273	TGTCCTGCCACGCGCTGCTGCTGCTGGCGCGGTGACAGTGGGCACCT	88224
204HisGlyGluAsn.....	207
88223	GCTGACGCGGTCTCTGCACCCACCTGCACGCGGAACGCCCGGCCAGTT	88174
208GlnAsnPheGlyLeuGlyLeut	215
88173	GAGGTCTCAGTGGCCTCGCTGTGACCTCACAACACGAGACTGGGT...C	88127
215	yraLaThrGly AlaIleHisLeuAspThrGlnGlyPheArgLysTrpGl	231
88126	ACAGGTTCAGTGTGCACCTCTCGACCCACACAGGC.....GG	88089
231	y.....AlaGlnPheSerGluThrAsnSerIleC	241
88088	CACCTGCCCAAGCAAGTAGTGTGCCAGACCTCGCTCTGAGCTCAGCT	88039
241	ys.....ArgHisValLeuPro	246
88038	GCAGCAGATGGCGCTCTCTGCTGCC	88013
seq_name: gb_pr:AP001053		
seq_documentation_block:		
LOCUS	AP001053	161920 bp DNA linear PRI 14-JAN-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 21, clone:KB836E9, MX1-D21S171 region, complete sequence.	
ACCESSION	AP001053	
VERSION	AP001053.1	GI:6693603
KEYWORDS	HTG	
SOURCE	Homo sapiens pre-pro-B cell cell_line:FLEB14-14 DNA, clone_lib:Keio BAC library clone:KB836E9.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.	
TITLE	1 (bases 1 to 161920)	
JOURNAL	Homo sapiens genomic DNA, chromosome 21, clone KB836E9, MX1-D21S171 region	
REFERENCE	Published Only in Database (2000) In press	
AUTHORS	Shimizu,N., Kudoh,J. and Shibuya,K.	
TITLE	2 (bases 1 to 161920)	
JOURNAL	Direct Submission	
REFERENCE	Submitted (12-JAN-2000) to the DBJ/JEMBL/GenBank databases.	
AUTHORS	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan	
TITLE	(E-mail:nshimizud@mb.med.keio.ac.jp, Tel:81-3-3351-2370, Fax:81-3-3351-2370)	
FEATURES	Location/Qualifiers	
source	1..161920	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/cell_line="FLEB14-14"	
	/cell_type="pre-pro-B cell"	
	/chromosome="21"	
	/clone="KB836E9"	
	/clone_lib="Keio BAC library"	
BASE COUNT	35486 a 41730 c 42915 g 41789 t	

204HisGlyGluAsn..... 207
58905 GGTAGCGGGTCTCTGAGCAGCACCTGCACGGGAGACGCCCGCCAGTT 58856
208GlnAsnPhGlyLeuGlyLeuT 215
58855 GAGGTCCTCCAGTGGCTCGTGTGACCTCAGAACACGAGGACTGGT...C 58809
215 yralaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGI 231
58808 ACAGGTCAGGTGTGCACCTCTCGACCCACACACAGC.....GG 58771
231 y.....AlaGlnPheSerGluThrAsnSerIleC 241
58770 CCAGCTGCCCAAGCAAGTGTGCCAGACCTCCGCTCGTGTGAGCTCAGCT 58721
241 ys.....ArgHisValLeuPro 246
58720 GCAGCAGATGGCGCTCTCTGTGTCGCC 58695
seq_name: gb_htg:AC079571
seq_documentation_block:
LOCUS AC079571
DEFINITION Mus musculus clone RP23-72P2, WORKING DRAFT SEQUENCE, 15 unordered
pieces.
ACCESSION AC079571
VERSION AC079571.1 GI:9964936
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 245130)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 245130)
DOE Joint Genome Institute.
Direct Submission
Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1764882
Center clone name: RPCI-23_72P2

Summary Statistics
Consensus quality: 233470 bases at least Q40
Consensus quality: 238342 bases at least Q30
Consensus quality: 239406 bases at least Q20
Estimated insert size: 224000; agarose-fp estimation
Quality coverage: 9.68 in Q20 bases; sum-of-contigs estimation
Quality coverage: 8.9 in Q20 bases; agarose-fp estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1669: contig of 1669 bp in length
* 1670 1769: gap of unknown length
* 1770 3529: contig of 1760 bp in length
* 3530 3629: gap of unknown length
* 3630 5387: contig of 1758 bp in length
* 5388 5487: gap of unknown length

Tue Sep 17 07:27:30 2002

55902 CCACAAAGAATCCAGCATGGGACTCAC.....TGGGGTC 55868
 181 roAspLeuGluIleLysSerClnAlaLeuTyrgluLeuGlnAsnArgLeu 197
 55867 CAAGCTCCAGCTCAAGCTCAGCTC.....TGG 55839
 198 CysGlnTyrrTpLeuGluHisGlygluAsnGlnAsnPheGlyLeuGlyLe 214
 55838 TGCCATCATGTTACAGCACTCCCAACCCCAACCTTTGTACTC....TG 55792
 214 uTyrrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg.LysTrp 230
 55791 TCACAGACCATTAAGCC...CACACTGATCTTACAGCAAGCAGCAAGAGT 55745
 231 GlyAlaGlnPheSer.....GluThrAsnSerIleCysArg 242
 55744 GGTACAGACATGTCGTGTGACGAGTGAAGACAAACAAGAAAGTGTCTGTGA 55695
 242 gHisVal 244
 55694 GCATGTA 55688

seq_name: gb_pr:AP001752

seq_documentation_block:
 LOCUS AP001752 340000 bp DNA linear PRI 30-MAY-2000
 DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 96/105.
 ACCESSION AP001752 AL163297 BA000005
 VERSION AP001752.1 GI:7768757

KEYWORDS
 SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (sites)
 Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
 Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
 Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
 The DNA sequencing consortium
 and sequencing consortium
 Nature 405 (6784), 311-319 (2000)
 20289799
 2 (bases 1 to 340000)
 Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
 Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
 Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
 Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
 Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
 Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,
 Hornischer,K., Barand,P., Scharfe,M., Schoen,O., Desario,A.,
 Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
 Hennig,S., Riesselmann,D., Dagand,E., Wehrmaeyer,S., Borzym,K.,
 Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
 Yaspo,M.L.
 Direct Submission
 Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
 Consortium. * RIKEN Genomic Sciences Center, Human Genome Research
 Group * Institute of Molecular Biotechnology, Genome Analysis *
 Keio University School of Medicine, Dept. of Molecular Biology *
 GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
 Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717416.
 The chromosome 21 mapping and sequencing consortium consisting of
 * RIKEN Genomic Sciences Center, Human Genome Research Group, *
 Sagamihara 228-8555, Japan,
 * e.mail: hattori@gsc.riken.go.jp
 * URL: http://hgp.gsc.riken.go.jp/
 and
 * Institute of Molecular Biotechnology, Genome Analysis, *

Beutenbergstrasse 11, D-07745 Jena, Germany,
 * e.mail: gscj-submit@genome.imb-jena.de
 * URL: http://genome.imb-jena.de/
 and
 * Keio University School of Medicine, Molecular Biology, * Tokyo
 160-8582, Japan,
 * e.mail: nshimizudmb-med.keio.ac.jp
 * URL: http://www.dmb.med.keio.ac.jp/
 and
 * GBF, Dept. of Genome Analysis,
 * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
 info.genome@gbf.de
 * URL: http://genome.gbf.de/
 and
 * Max-Planck Institute for Molecular Genetics,
 * Ihnestrasse 73, D-14195 Berlin, Germany,
 * e.mail: info-chr21@molgen.mpg.de
 * URL: http://chr21.rz-berlin.mpg.de/
 AL163297: Submitted (10-Apr-2000).

FEATURES

source
 1. .340000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 <1. .39630
 /note="Accession No. AP001051"
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.3"
 /clone="KB161B12, 5' partial"
 /clone.lib="Keio BAC library"
 924. .2304
 /note="LIM1"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 2307. .2592
 /note="AluJo"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 2593. .2616
 /note="(CAAAA)n"
 /rpt_family="Simple_repeat"
 /rpt_type=TANDEM
 2617. .3063
 /note="L1PA16"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 3712. .3856
 /note="L1MC/D"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 3927. .4224
 /note="AluJb"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 4285. .4400
 /note="L1MC/D"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 4657. .4874
 /note="AluJb"
 /rpt_family="SINE/Alu"
 /rpt_type=DISPERSED
 4968. .5280
 /note="LINE"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 5798. .5987
 /note="L1MD"
 /rpt_family="LINE/L1"
 /rpt_type=DISPERSED
 5989. .6291

```

/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 6302..6441
/note="FLAM_A"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 6530..7150
/note="MLT12"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
mRNA complement(Join(<7441..7566,23941..24073,26891..27040,
37908..38011,50206..50356,51680..51939,53005..53112))
/note="HSF2Bp"
/note="Accession No. AB007131"
/product="heat shock transcription factor 2 binding
protein"
gene complement(7441..53112)
/note="HSF2Bp"
CDS complement(Join(<7441..7566,23941..24073,26891..27040,
37908..38011,50206..50356,51680..51715))
/note="HSF2Bp"
/note="Accession No. AB007131"
partial"
/codon_start=1
/product="heat shock transcription factor 2 binding
protein"
/protein_id="BAA95539.1"
/db_xref="GI:7768758"
/translation="MGEAGAAEEACRHMGTKEEFVKKKDLRLTTEVMQIRDFLPR
LLNGVEVQFKLIVKRLERKEQELKMDCEHFKARLETVOADNIREKKKELAL
RQNLNAKQOLLOQAYCYETMGAACTLLWGVSSSEVVKAILGDKALKAFPSITGOT
MESFVKSLDGDOELDSQSFVVALGIVTNVAALACGREFLVNSSLVLLDTILQLL
GOLKPGQCTKLKVG"
exon complement(7441..7566)
/note="HSF2Bp"
/number=7
repeat_region 7702..7999
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 8247..8455
/note="MER46A"
/rpt_family="DNA/MER2_type"
/rpt_type=DISPERSED
repeat_region 9535..9640
/note="MLT1G"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 9714..9959
/note="MLT1F"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 9981..10277
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 10278..10575
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 10708..10829
/note="MLT1H"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 11291..11591
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region 11596..11892
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED

```

```

repeat_region 12004..12273
/note="LIM4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 12274..12590
/note="AluY"

alignment_scores:
  Quality: 102.00      Length: 326
  Ratio: 0.823         Gaps: 20
  Percent Similarity: 38.037   Percent Identity: 23.926
alignment_block:
US-09-674-779-2 x AP001752/rev ..

Align seg 1/1 to reverse of: AP001752 from: 1 to: 340000

16 MetLeuValAlaCysSerAlaProIleProThrAsnProGlnValSerPr 32
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211329 ATGCTGCTAGCGTGCACCTGGCAGCATCTCTACGTACCGAAACAGCAG 211280
32 oileLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH 49
:::
211279 GCTA.....CACT 211272

49 isThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis 65
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211271 CCTGCTCTGATCACAACACACTGCTCGCGAGGGG.....CGG 211231

66 PheGluThrTrpLeuGlnMetHisAlaThrLysGlnGluValValar 82
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
211230 ACACAATGCTGGCTGCAGCTGACGACGCCACCG..... 211198

82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetS 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211197 .TACGAGCGCTTG.....CCGAGACTGGGATCTGCTGTGACACTATGTGGGAGGGCG 211155

99 erGlnLeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGlu... 114
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
211154 AGCCCAAGGTCGTTGGGCTGTCTGTGACACTATGTGGGAGGGCG 211105

115 ProTyrGlnLeuPro.....ProGluHisLe 123
||| |||||
211104 CCAGGCTGGCTACCCAGTCTAGGTTGGGACCACAGCACTGAGGAGAGCC 211055

123 uTTPGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerA 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211054 CTGGGGA.....TGCCCACTGAGCCTCTCTGTGTACAGCCCTCAGAAGCC 211011

140 rGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr..... 153
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
211010 AC.....TGCCCACTAACACACAGCCACCCGGTTTCTGGGGCGGAGCC 210967

154 .....ArgAsnProGluLeuAsnGln... 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210966 AGGCTGGGGCCAGTCCCTCTCTGACCGCGCACCCAGAGGCATGCAAG 210917

160 ..... 160

210916 AGGCCCCAGGGCTGGGCCCCAGGAGAAACACTGACCTCAGAGCCAGAG 210867

161 .....CysAlaGlyAlaAlaMetSerLysHisLeuThrAsnSer 174
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210866 GAGGCGCTTTGTGTGGGAGGAGTGCCTGAGCTGCTTCGAGCCTCGAAA 210817

175 AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeu... 190
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
210816 GGTGTACCTCTGG.....GGCGTCAGACACAGGAGGAGCC 210779

191 .....TyrGluLeuGlnAsnArgLeuCysGln. 199
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
210778 AGTGTGCTAAATGTCTTCCCTGGCCCTCCAGCATCTCCCTGTGTC 210729

```

[illegible]

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (unpublished.) for Human and Mouse sequences.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. description of the metrics can be found

117744	AAACACTGAGTCTTGATATAAACACTTTGAAAAATCTCCTTCACGTT	117793
200	yrTrpLeuGluHisGlyGluAsn...GlnAsnPheCysLeuGlyLeuTy	215
117794	GTCACATCAACACCATACCTCTGACATCAACACTCGCGTGGGAACAGC	117843
216	AlaThrClyAlaIleHisLeuAspThrGlnGlnPheArgLysTrpClyAl	232
117844	GCA.....C	117869
232	agInPheSerGluThrAsnSerIleCysArgHisValLeuProLysAsnL	249
117870	ATTATACCTCGGTTCTCTCCTCAGATGTATACATCTTTGCCTTTACTA	117919
249	ys 249	
117920	AA 117921	
seq_name: gb_pr:AL133414		
seq_documentation_block:		
LOCUS	AL133414	159169 bp DNA linear PRI 01-JUN-2001
DEFINITION	Human DNA sequence from clone RP5-1060P11 on chromosome 19.	
	Contains ESTs, STSS and GSSs. Contains 10 natural killer cell	
	receptors, including the KIR2DL2, KIR2DL4, KIR2DS1 and KIR2DS2	
	genes for killer cell two domain long cytoplasmic tail	
	immunoglobulin-like receptor 2 and 4 and short cytoplasmic tail	
	immunoglobulin-like receptor 1 and 2 and the KIR3DL2 and KIR3DS1	
	genes for killer cell three domain short cytoplasmic tail	
	immunoglobulin-like receptor 2 and short cytoplasmic tail	
	immunoglobulin-like receptor 1, complete sequence.	
ACCESSION	AL133414	
VERSION	AL133414.4	GI:14329489
KEYWORDS	HTG; KIR2DL2; KIR2DL4; KIR2DL5; KIR2DS1; KIR2DS2; KIR3DL2; KIR3DS1;	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 159169)	
JOURNAL	Milne, S.	
COMMENT	Direct Submission Submitted (26-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequests@sanger.ac.uk On Jun 8, 2001 this sequence version replaced gi:7242326. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated as part of a collaboration with Michael Wilson, Michaela Torker and John Trowsdale, Department of Pathology, Division of Immunology, the University of Cambridge, UK. RP5-1060P11 is from the library RPI-5 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: pCYPAC2 This sequence is the entire insert of clone RP5-1060P11 The true right end of clone RPI-52N12 is at 11343 in this sequence.	


```

DEFINITION Homo sapiens BAC clone RP11-632K21 from 7, complete sequence.
ACCESSION AC069304
VERSION AC069304.7 GI:13549358
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 161247)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 161247)
AUTHORS Isak, A., Maupin, R. and Reed, J.
TITLE The sequence of Homo sapiens BAC clone RP11-632K21
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 161247)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 4 (bases 1 to 161247)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 5 (bases 1 to 161247)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 161247)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Apr 5, 2001 this sequence version replaced gi:13431246.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0632K21
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc
SOURCE INFORMATION:

```

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Taten, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP5-1051J4, 200 bp overlap.
Actual start of this clone is at base position 152813 of RP5-1051J4; actual end is at base position 161247 of RP11-632K21.

FEATURES

```

Location/Qualifiers
1..161247
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-632K21"
/clone_lib="RPCI-11"
repeat_region 40..71
/rpt_family="(TAA)n"
repeat_region 2090..2147
/rpt_family="(GAAA)n"
repeat_region 2280..2340
/rpt_family="MaLR"
misc_feature 2381..2572
/note="similar to EST BG033953 (NID:gl2426761)"
misc_feature 2569..2785
/note="similar to EST W67711 (NID:gl376582) zd42e02.rl"
misc_feature 2595..2633
/note="similar to EST AW007542 (NID:g5856405) wt02b11.x1"
misc_feature 2595..2633
/note="similar to EST AL518924 (NID:gl2782417)"
misc_feature 2649..2785
/note="similar to EST AL518924 (NID:gl2782417)"
misc_feature 2652..2785
/note="similar to EST AW007542 (NID:g5856405) wt02b11.x1"
repeat_region 2792..2896
/rpt_family="L1"
misc_feature 2887..3237
/note="similar to EST W67711 (NID:gl376582) zd42e02.rl"
misc_feature 2887..3280
/note="similar to EST AL518924 (NID:gl2782417)"
misc_feature 2887..3280
/note="similar to EST AW007542 (NID:g5856405) wt02b11.x1"
misc_feature 2899..3275
/note="similar to EST BG403336 (NID:gl3296784)"
misc_feature 2904..3275
/note="similar to EST BF541175 (NID:gl1628556)"
misc_feature 2961..3275
/note="similar to EST BF982899 (NID:gl2385711)"
repeat_region 4160..4427
/rpt_family="L1"
repeat_region 4429..4832
/rpt_family="L1"
repeat_region 4846..4966
/rpt_family="L2"
misc_feature 5048..5290
/note="similar to EST AT002182 (NID:g3202516) oq85c05.s1"
repeat_region 5285..6565
/rpt_family="L1"
repeat_region 7323..8249
/rpt_family="L1"
repeat_region 9458..9606
/rpt_family="MaLR"
repeat_region 9618..9971
/rpt_family="L1"
repeat_region 9975..10442
/rpt_family="L1"
repeat_region 10523..10828
/rpt_family="ERV1"

```


DAVYSEAMVGGRLTQDLISDWMALNQNGTTCGRSVKAPRVVAVNRMLSKREMSASATSA
SEKPGAVAISSRSFSPVKMH"
complement(2152. .3693)
/gene="Atu3764"
CDS
complement(2152. .3693)
/gene="Atu3764"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="sigma 54 dependent transcriptional activator"
/protein_id="AA144574.1"
/db_xref="GI:17742191"
/translation="MTAHLVLIIDDDPVRRLKNAVERXGHLALLAENGRAGLELLKQ
YSGINVLIDLAMPDMGDLAFKAVGELGTDPVIVQGGGIDTVQVOMRAGAFDF
VWVPSPERIGALSNAKLKDRKEAKARTGRGRSNAFTDIVSASPAMLRVIELAQ
RAQSSIPVVESESGVGKEMVARATQSGDRANKPFTVNCGAIPHNLVESILFGHE
KGATGATGRHKVFGADGGTFLDEIGDLPDQVQKLLRAVQOGEIETVGSARVOK
VNRLLSATNKLIEEYKGRFREDYLRNLNVPFITPALRRKEDIPLHARVAFER
SNAOKLPNPVGLDASALALLATYDFGNIRQLENVAFPRVLLSQSGSELSDPPQAL
QLPFGSGNIGDDDDIIRSPSGSLKLAAYSPPAASHAVIEEPDSITTYIRGNLIS
STDPSGNIRKLAETIEELIRFALRFYRQMSQVARKLIGRSTLYRKLDYGDIDPDP
LRDVA"
3863. .5713
/gene="pepf"
/note="Atu3765"
3863. .5713
/gene="pepf"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="oligoendopeptidase F"
/protein_id="AAL44575.1"
/db_xref="GI:17742192"
/translation="WTLRRAPHPVFLAETSGPVLGDLPGWKLSDLYPSGDSPEYKR
DLEKASMAAFERAKKLESAKAKAGEDIGALADYALDLDLIGRIGSPAGLIYF
SDTNPANGKLDGDAQAKLEISAHLLFTLELNRIDDVADMACMADAAGAARPLWL
VDLRKPQHLDDLEQLEFLKSMTSAAANRLNFDMDADLDFIDGASLSLEVTLMN
LOEPEPTRKAAEALSAFKNLNRVFTLITLAKDKDISDRWNPEDTADSRHLAN
RVERVDVLAAYAKAVPRLSHRYAMKAKWLQMEQWYDRNAPLPTDNTIIPWD
EAKDTVLSAVGNFAPENADTARFFDEWDIPARCPKAPCAFAHPTVPSAHPVVLVN
YLGPRVMTLAHELGHQVLAGOQALMCATPLLAETASVGEMLFERKLDDT
KDIREKARKVVEDMINTVROIATFERKLHARKKEGELTSEIGELSLVQAE
SLGPAIRISEGYETWYIHPFIHSFVYVAYAFGDCVLNLAAYOKAAGFOEYF
ELLKAGGTKHSELLKPFGLDTPDFWDKLSMIEGLIDELELDKQA"
5813. .6448
/gene="Atu3766"
5813. .6448
/gene="Atu3766"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL44576.1"
/db_xref="GI:17742193"
/translation="MKRPADLTTLRETLRWEPQTFQRIQDHRRLRSADALFRAP
NVAVLKQYGGKPLCVRLVMNKGEGLDIETVEFRPLPENAVVRVIRARQTTLDSA
DTFVHKSSRREPVDAAAEFSAEADEVLLNERGELCEGSSIFVEMWGPMLTP
PLDSGLFGLRDLRRRRRQALRPEDIAGRKIFVGNLSLEGLIAELI"
complement(6445. .7353)
/gene="Atu3767"
complement(6445. .7353)
/gene="Atu3767"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="GDEF family protein"
/protein_id="AAL44577.1"
/db_xref="GI:17742194"
/translation="MNDLLERVAALQSATPVLLIAYDSNDRLRHANSFRESFLADG

DFPTWADIMLRNLHRLRTGVVIRNDFEGWLLSAQSRGKLPYRAFEMDLVDGRFWMT
ETVQDGMCLVCASDVTELDQFYRTLRQDRDLAVRASOTDELGTGTAARRVEFAKLAQ
VEQYSGKLEAPCLDITDFKQINDRFHGKCDGAVLDRDFVIRRVDCFGRI
GGEFMILIPRSSIAEGREIVERLQKVRSLRPLSEPFDFSYSCSAGMTVFRMMENAD
DYFMRADOALYAAKEKGRDRLLAVT"
7613. .9058
/gene="Atu3768"
7613. .9058
/gene="Atu3768"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="homospermidine synthase"
/protein_id="AAL44578.1"
/db_xref="GI:17742195"
/translation="MTDANYPIYGDIDGPIIMIGFSGISGRGLPLIERHFNDFRRLV
VIDPREDTAAFLAERNIRHVOTHTKENYKDVLPKLLKGVQGVKCNLSYDAASVDL
MRICKHMLYDITVVEPVGFTDADADNARTNYSURETLMKERKPGGTAVST
CGANPGWVFWKQALVNLAKDTGLDIEPQSHDRGKAKLMQOLGVKGVHVAERDQ
RAKEPKPFGAFTWNTSVEGFIAEGFPQPAELGWHENMMPKNAKKQKKGSKAALYLDQ
PGANTRVTCPLGAQYGFVLTNEAISISDIETVRNDDGDVITYRPTCHTAYHPCND
ATLSHELFGNGGNPQPIQHVLGDELVDGDELGVLLYHGRNAYWYSGLSLEEAR
SLAPSNATGLQVSSAVLAGMAWAIENPEAGIVEADEMDYRRCLVEQRYLGPVEGHI
TDWTPLDGRPGGLFPDNDIMSDPWFQFNILVR"
9208. .9582
/gene="Atu3769"
9208. .9582
/gene="Atu3769"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="outer membrane lipoprotein"
/protein_id="AAL44579.1"
/db_xref="GI:17742196"
/translation="MNFKTSALLSAAIAVSSCVAPGPSQOTSMAPSLSRGPVADGRW
IDRNGIISTFQNGAFSTRSDNTILLASGVTVISPTLYEINMTSLVRNTQSRVNCAL
VSPSQLNCTTDANSQFSLTRQG"
9798. .11708
/gene="Atu3770"
9798. .11708
/gene="Atu3770"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="5'-nucleotidase"
/protein_id="AAL44580.1"
/db_xref="GI:17742197"
/translation="MKKILGLGLSVATLSAGAAALADYELNLIHINDEHSRIESIN
KFDSTCSAEEGKNECFGGAARLLTAINGOTDALKAGCKNVLLNAGDNFQGLFTTT
YKGTVEAEVILNMFADMTVGNHEFDDSDGLAGFLDKVQFVTVVAVATAASKIGD
RVKPSITLEVGGKIGIGAVANDTAELATPGNITITADIAEKIGEQVOKLQDQGVNK
IIALTHVGPDLDEFTAKIPDVVVVGGHSHTLLSNTDQKAEQPTLVDPNGYKVP
VQAGQYSKYLDGVFDGDKVVKESGDPILIDSTFKPDGATLKRIDEIKAPTEAL
KSKVGTSEPIEGDKVCRKESGMNLVADATLARKVQDQGTITAFANSGLGRSID
GGDVSMEGVLTVLPQNTVATFQKEDIRALENGSQIDDVAGREWQVQSKYSED
RKSPGARSIVSDVREKGFVPLDPTKTYTVAANNVFRVGGGCFKVFATNAINAYDFG
PNLEDAVAYITANSKYPTDSEVTAGYVAPAKPAPAPAPAPAPAPAPAPAPAT
APAPATAPAT

alignment_scores: Quality: 101.00 Length: 169
Ratio: 1.174 Gaps: 9
Percent Similarity: 50.888 Percent Identity: 22.485
alignment_block: US-09-674-779-2 x AE009307/rev
Align seg 1/1 to reverse of: AE009307 from: 1 to: 11913
74 HisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerAr 90

Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, D., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.	
Direct Submission	
Unpublished	
2 (bases 1 to 194508)	
Worley, K.C.	
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
----- Genome Center	
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	
----- Project Information	
Center project name: GLGG	
Center clone name: CH230-50L12	
----- Summary Statistics	
Assembly program: Phrap; version 0.990329first call to findhaplolist	
Consensus quality: 169488 bases at least Q40	
Consensus quality: 176660 bases at least Q30	
Consensus quality: 183878 bases at least Q20	
Estimated insert size: 174717; sum-of-contigs estimation	
Quality coverage: 0x in Q20 bases; agarose-fp estimation	
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation	

* NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).	
* NOTE: This is a 'working draft' sequence. It currently consists of 59 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
*	1 11545: contig of 11545 bp in length
*	11546 11645: contig of unknown length
*	11646 19125: contig of 7480 bp in length
*	19126 19225: gap of unknown length
*	19226 27540: contig of 8315 bp in length
*	27541 27640: gap of unknown length
*	27641 34907: contig of 7267 bp in length
*	34908 35007: gap of unknown length
*	35008 40495: contig of 5488 bp in length
*	40496 40595: gap of unknown length
*	40596 42662: contig of 4667 bp in length
*	42663 45362: gap of unknown length
*	45363 51446: contig of 6084 bp in length
*	51447 51546: gap of unknown length
*	51547 56966: contig of 5420 bp in length
*	56967 57066: gap of unknown length
*	57067 61760: contig of 4694 bp in length
*	61761 61860: gap of unknown length
*	61861 66676: contig of 4816 bp in length
*	66677 66776: gap of unknown length
*	66777 70702: contig of 3926 bp in length
*	70703 70802: gap of unknown length
*	70803 75586: contig of 4784 bp in length
*	75587 81828: contig of 6142 bp in length
*	81829 86907: gap of unknown length
*	86908 87007: contig of 4979 bp in length
*	87008 90055: gap of unknown length
*	90056 90155: contig of 3048 bp in length
*	90156 95728: contig of 5573 bp in length
*	

ORGANISM

Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 4383)
Kraev, A.S.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

Location/Qualifiers
1. 4383
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="yk49f8 (765-4383)"
/clone_lib="Yuji Kohara CDNA"
/dev_stage="mixed"
1..22
/function="trans-spliced leader SL1"
/evidence="experimental"
84..3770
/gene="mca-1"
84..3770
/gene="mca-1"
/codon_start=1
/product="calcium ATPase"
/protein_id="CAA11491.1"
/db_xref="GI:2826864"
/db_xref="SPTREMBL:O45215"
/translation="MRLRRGTYDEDAKEFCGLDRLGLMEARGAEAIIVRLSTEHEG
VGLCKLKTSLVGEQADLRHRYVANTIPPAKSGFVRLVLDACKDPTLVI
LVLGFLINLALSYFEPTSAADATOLHVNATTAIANGTFMSTTEAPSEGHGTAWIE
GVAILLVVLVAVNDYSKEROFRSLQEKIFGQKFSVIRNGEADVPVSLVVG
DIARYKGDLLPADGLIQSNDLKIDESSLIGESDHIKKIESDPVLLSGYAMEGSG
KMLITAVNSOTGIIMLLGKAGIGDDSTSSSSSSSSSSSSSSSSSSSSSSSSSSSS
SGDDTLAKVLAQKSLAQIYYCGTTIAIILVLTFRCLDHYVFEKNEFSLVLD
IQMFYKFTIAVITLIVISPEGLPLAIALALTYSVKMMHNNLVRHLDACETMGNA
STCSKTGTLTNNRVQSYINGNHYTSQAQPHGANLPGSTGPIILMEALSVCAYN
SMIVPTKAGEIOQLGNKTEGLLGFVNRLLGGDYAAIRKFPEDHLLKVVTFNSRK
CMNTVVPVANGONIGYRVCKGASEIVLGRCTVLIGSDGKPHQLTCDRLKEITSTII
HEMANSGLTICVAKTIKKGRDVEKTEIEFAEDSDIDWDDADAMQNTGTGATG
IQDPVRPEVPAISKCKKAGITVRMTGDNIIMTAKIAMSKLILEPDEDFLALGKEF
NERIDENGKVSQAKDEIPRLRLVLAQAQADKTYLLVKGIDSKATPQREIVATG
GTNDGPKLKADGVGFAMGTADTVAKEASDIILTDNFTSVKAVMMGRNVDSTISF
LQFOLTVNVAVITAFVGAIVTSDSPKAVHMLINLIMDTLASLALATEOPTDELLE
RPYGRKSLISRTVWKNLTCHALYQLIIIEVIEFGDTIFGKTGLYAPLFAFAPSOH
FTLVNAPVMTVFEINARKVHGERNVFKGLASNRVFCVIVWTFIAQIILVDFGGA
WFSTAPLTLQNVICVLVGLSTLIWQIVATIPSKLLPKAWKVGKEVOPANLHNGD
YXNRASQVTLRRSGKSLWRGMFIIIGNHLRLVRAFMEKSEKAAFGRTAPAMTAA
AERWASRYKRYHQHQKQKATAETAESVKSADWAKEQKQKKTFFQIKQVARGKSLD
KDSKHHKRRKQDQVNDMEDIELN"

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

Location/Qualifiers
1. 4383
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="yk49f8 (765-4383)"
/clone_lib="Yuji Kohara CDNA"
/dev_stage="mixed"
1..22
/function="trans-spliced leader SL1"
/evidence="experimental"
84..3770
/gene="mca-1"
84..3770
/gene="mca-1"
/codon_start=1
/product="calcium ATPase"
/protein_id="CAA11491.1"
/db_xref="GI:2826864"
/db_xref="SPTREMBL:O45215"
/translation="MRLRRGTYDEDAKEFCGLDRLGLMEARGAEAIIVRLSTEHEG
VGLCKLKTSLVGEQADLRHRYVANTIPPAKSGFVRLVLDACKDPTLVI
LVLGFLINLALSYFEPTSAADATOLHVNATTAIANGTFMSTTEAPSEGHGTAWIE
GVAILLVVLVAVNDYSKEROFRSLQEKIFGQKFSVIRNGEADVPVSLVVG
DIARYKGDLLPADGLIQSNDLKIDESSLIGESDHIKKIESDPVLLSGYAMEGSG
KMLITAVNSOTGIIMLLGKAGIGDDSTSSSSSSSSSSSSSSSSSSSSSSSSSSSS
SGDDTLAKVLAQKSLAQIYYCGTTIAIILVLTFRCLDHYVFEKNEFSLVLD
IQMFYKFTIAVITLIVISPEGLPLAIALALTYSVKMMHNNLVRHLDACETMGNA
STCSKTGTLTNNRVQSYINGNHYTSQAQPHGANLPGSTGPIILMEALSVCAYN
SMIVPTKAGEIOQLGNKTEGLLGFVNRLLGGDYAAIRKFPEDHLLKVVTFNSRK
CMNTVVPVANGONIGYRVCKGASEIVLGRCTVLIGSDGKPHQLTCDRLKEITSTII
HEMANSGLTICVAKTIKKGRDVEKTEIEFAEDSDIDWDDADAMQNTGTGATG
IQDPVRPEVPAISKCKKAGITVRMTGDNIIMTAKIAMSKLILEPDEDFLALGKEF
NERIDENGKVSQAKDEIPRLRLVLAQAQADKTYLLVKGIDSKATPQREIVATG
GTNDGPKLKADGVGFAMGTADTVAKEASDIILTDNFTSVKAVMMGRNVDSTISF
LQFOLTVNVAVITAFVGAIVTSDSPKAVHMLINLIMDTLASLALATEOPTDELLE
RPYGRKSLISRTVWKNLTCHALYQLIIIEVIEFGDTIFGKTGLYAPLFAFAPSOH
FTLVNAPVMTVFEINARKVHGERNVFKGLASNRVFCVIVWTFIAQIILVDFGGA
WFSTAPLTLQNVICVLVGLSTLIWQIVATIPSKLLPKAWKVGKEVOPANLHNGD
YXNRASQVTLRRSGKSLWRGMFIIIGNHLRLVRAFMEKSEKAAFGRTAPAMTAA
AERWASRYKRYHQHQKQKATAETAESVKSADWAKEQKQKKTFFQIKQVARGKSLD
KDSKHHKRRKQDQVNDMEDIELN"

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

Location/Qualifiers
1. 4383
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="yk49f8 (765-4383)"
/clone_lib="Yuji Kohara CDNA"
/dev_stage="mixed"
1..22
/function="trans-spliced leader SL1"
/evidence="experimental"
84..3770
/gene="mca-1"
84..3770
/gene="mca-1"
/codon_start=1
/product="calcium ATPase"
/protein_id="CAA11491.1"
/db_xref="GI:2826864"
/db_xref="SPTREMBL:O45215"
/translation="MRLRRGTYDEDAKEFCGLDRLGLMEARGAEAIIVRLSTEHEG
VGLCKLKTSLVGEQADLRHRYVANTIPPAKSGFVRLVLDACKDPTLVI
LVLGFLINLALSYFEPTSAADATOLHVNATTAIANGTFMSTTEAPSEGHGTAWIE
GVAILLVVLVAVNDYSKEROFRSLQEKIFGQKFSVIRNGEADVPVSLVVG
DIARYKGDLLPADGLIQSNDLKIDESSLIGESDHIKKIESDPVLLSGYAMEGSG
KMLITAVNSOTGIIMLLGKAGIGDDSTSSSSSSSSSSSSSSSSSSSSSSSSSSSS
SGDDTLAKVLAQKSLAQIYYCGTTIAIILVLTFRCLDHYVFEKNEFSLVLD
IQMFYKFTIAVITLIVISPEGLPLAIALALTYSVKMMHNNLVRHLDACETMGNA
STCSKTGTLTNNRVQSYINGNHYTSQAQPHGANLPGSTGPIILMEALSVCAYN
SMIVPTKAGEIOQLGNKTEGLLGFVNRLLGGDYAAIRKFPEDHLLKVVTFNSRK
CMNTVVPVANGONIGYRVCKGASEIVLGRCTVLIGSDGKPHQLTCDRLKEITSTII
HEMANSGLTICVAKTIKKGRDVEKTEIEFAEDSDIDWDDADAMQNTGTGATG
IQDPVRPEVPAISKCKKAGITVRMTGDNIIMTAKIAMSKLILEPDEDFLALGKEF
NERIDENGKVSQAKDEIPRLRLVLAQAQADKTYLLVKGIDSKATPQREIVATG
GTNDGPKLKADGVGFAMGTADTVAKEASDIILTDNFTSVKAVMMGRNVDSTISF
LQFOLTVNVAVITAFVGAIVTSDSPKAVHMLINLIMDTLASLALATEOPTDELLE
RPYGRKSLISRTVWKNLTCHALYQLIIIEVIEFGDTIFGKTGLYAPLFAFAPSOH
FTLVNAPVMTVFEINARKVHGERNVFKGLASNRVFCVIVWTFIAQIILVDFGGA
WFSTAPLTLQNVICVLVGLSTLIWQIVATIPSKLLPKAWKVGKEVOPANLHNGD
YXNRASQVTLRRSGKSLWRGMFIIIGNHLRLVRAFMEKSEKAAFGRTAPAMTAA
AERWASRYKRYHQHQKQKATAETAESVKSADWAKEQKQKKTFFQIKQVARGKSLD
KDSKHHKRRKQDQVNDMEDIELN"

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

FEATURES

Location/Qualifiers
1. 4383
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="yk49f8 (765-4383)"
/clone_lib="Yuji Kohara CDNA"
/dev_stage="mixed"
1..22
/function="trans-spliced leader SL1"
/evidence="experimental"
84..3770
/gene="mca-1"
84..3770
/gene="mca-1"
/codon_start=1
/product="calcium ATPase"
/protein_id="CAA11491.1"
/db_xref="GI:2826864"
/db_xref="SPTREMBL:O45215"
/translation="MRLRRGTYDEDAKEFCGLDRLGLMEARGAEAIIVRLSTEHEG
VGLCKLKTSLVGEQADLRHRYVANTIPPAKSGFVRLVLDACKDPTLVI
LVLGFLINLALSYFEPTSAADATOLHVNATTAIANGTFMSTTEAPSEGHGTAWIE
GVAILLVVLVAVNDYSKEROFRSLQEKIFGQKFSVIRNGEADVPVSLVVG
DIARYKGDLLPADGLIQSNDLKIDESSLIGESDHIKKIESDPVLLSGYAMEGSG
KMLITAVNSOTGIIMLLGKAGIGDDSTSSSSSSSSSSSSSSSSSSSSSSSSSSSS
SGDDTLAKVLAQKSLAQIYYCGTTIAIILVLTFRCLDHYVFEKNEFSLVLD
IQMFYKFTIAVITLIVISPEGLPLAIALALTYSVKMMHNNLVRHLDACETMGNA
STCSKTGTLTNNRVQSYINGNHYTSQAQPHGANLPGSTGPIILMEALSVCAYN
SMIVPTKAGEIOQLGNKTEGLLGFVNRLLGGDYAAIRKFPEDHLLKVVTFNSRK
CMNTVVPVANGONIGYRVCKGASEIVLGRCTVLIGSDGKPHQLTCDRLKEITSTII
HEM

/rpt_family="MERGB"

alignment_scores:

Quality: 100.00 Length: 201
 Ratio: 1.136 Gaps: 9
 Percent similarity: 43.781 Percent Identity: 19.900

alignment_block:

US-09-674-779-2 x AC083964 ..

Align seg 1/1 to: AC083964 from: 1 to: 161160

```

23 ProfileProthrAsn..... 27
||||:|||||:
139277 CCGGTGCCACCTCAGCAGCCACACACGGAATCTAGAGCCATGCATGC 139326
28 .....ProGlnValSerProIleLysThrPro..... 36
|||||:||||:||||:
139327 GAACACGTGCCACCTCAGCAGCCACACACGGAATCTCAGAGCCAC 139376
37 .....SerValLeuIleThrLysAspLysIleGlyAsp 47
|||||:||||:||||:
139377 CTGGAACCAATGCCACTCAGCAGCCACACACGGAATCTCAGAGCC 139426
48 His.....HisThrHisGluHisAspGluSerVa 57
|||||:||||:||||:
139427 CATCAGAACCTGTACCCACTCAGCAGCCACACACGGAATCTCAGAG 139476
57 lserHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHis 74
|||||:||||:||||:
139477 CACACACTGGAAACCAATGCCACTCAGCAGCCACACACGGAATCT 139517
74 isAlaThrLysGlnGluValArgTyrGlnAlaTyrLeuGlnSerArg 90
|||||:||||:||||:
139518 ACAGAACTCAGCGCCACATCAGCAGCCACACACGGAATCTCAGAG 139567
91 LeuGlyAsnTyrLeuProMetSerGlnLeuLeuThrAlaArgSe 107
|||||:||||:||||:
139568 ACATGACACCGGTGCCACTCAGCAGCCACACACGGAATCTCAGAG 139617
107 r.....TrpGlnAlaCysGlyHisGluProTyrGlnLeuProGluH 122
|||||:||||:||||:
139618 CACACACTGGAACCCATGC.....CCACTCAGC 139646
122 is.....LeuTrpGlyGlnIleValProThr..... 130
|||||:||||:||||:
139647 ACAGTCAATATGGAATCAAGACACACTGGAACCAATGCCACCTC 139696
131 .....LeuHisLeuTyr 134
139697 TGCACACACACAGGAATCCAGAGCCACACAGGTAACCCACCCCA 139746
134 rGlnAspLeuLysSerArgGlyLe.....LeuP 144
|||||:||||:||||:
139747 CCTCAGCAGCCACCAACATGGAATCCAGAGCCACACCAATGGAAC 139796
144 rGlnAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGln 160
|||||:||||:||||:
139797 CCACCTCAGCAGCAAGTCAATGGAATCAAGAGCCACACACTGGAAC 139846
161 Cys 161
|||||
139847 TGC 139849

```

seq_name: gb_htr:AC090135

seq_documentation_block:

LOCUS AC090135 190173 bp DNA linear HTG 19-JAN-2002
 DEFINITION Homo sapiens chromosome 8 clone RP11-18D5 map 8, *** SEQUENCING IN
 PROGRESS ***, 4 unordered pieces.
 ACCESSION AC090135
 VERSION AC090135.9 GI:18250010
 KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFTN.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE


```

65 speGluThrTrp.....LeuGlnMethHisA 75
||||:||||:||||:
20507 TATTGAAGTTGGCGCAACGCTACCGAGGATCTCTCGGCATCTATC. 20858
||||:||||:||||:
75 laThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeu 91
|||:||||:||||:
20857 TAGCGGCTCGAGTAGTCTGACCGCTCAGACTCTCTGGCGAGAGCGGTTTC 20808
||||:||||:||||:
92 GlyAsnTyrLeuProMetSerGlnLeuLeuThrAlaArgSerTr 108
||||:||||:||||:
20807 AATTACCGCTCCAAATCTC...AAGTCATAGTTGCCGAGAAGAAGG 20761
||||:||||:||||:
108 pGlnAlaCysGlyHisGluProTyrGlnLeuProGluHisLeuTrpG 125
||||:||||:||||:
20760 CCAAGCGCTGGGCTGCTCTGCTCTCTAGCAATGCGGAGGACCGTTGG 20711
||||:||||:||||:
125 lGlnlleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGly 141
|||:||||:||||:
20710 GTACCCCTCGTCGACAAATCTCATGTTTGCACGCGGAGGCGGCGG 20661
||||:||||:||||:
142 IleLeuProAlaAsnThrGlnlleArgSerValTyrArgAsnProGluLe 158
|||:||||:||||:
20660 GTC.....GGAAGGCACTCT 20647
||||:||||:||||:
158 uAsnGlnCysAlaGlyAlaAlaMetSerLysHisLeuThrAsnSera 175
|||:||||:||||:
20646 TATTGGGTTGTCGCGGCTGGTCA...GCCGAAATATACCGGCGGCTGG 20600
||||:||||:||||:
175 laIleAspIleTrpValProAspLeuGluLeuLysSerGlnAlaLeuTyr 191
||||:||||:||||:
20599 GGTCGATCATGGTCTAGGAGTTAATGCCCTCGCGCGCGCTTCTAC 20550
||||:||||:||||:
192 Glu..... 192
||||:||||:||||:
20549 GAGAGATGGTGGTCAGGTAGTCGCGCGCTACCGAGCTCAACCTGA 20500
||||:||||:||||:
193 .....LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGly 205
||||:||||:||||:
20499 CGGAAGAATCCAGCTCGAAGTGTCTACTATTGGCTGCTTCTGGC 20454
||||:||||:||||:
seq_name: gb_htg.AC105039

```

```

seq_documentation_block:
LOCUS AC105039 145667 bp DNA linear HTG 10-FEB-2002
DEFINITION Homo sapiens chromosome 18 clone CTD-2357J19 map 18, WORKING DRAFT
SEQUENCE, 4 ordered pieces.
ACCESSION AC105039
VERSION AC105039.3 GI:18643649
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145667)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145667)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamasares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

```

```

Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 10, 2002 this sequence version replaced gi:18598788.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23024
Center clone name: 2357_J19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 145044 bases at least Q40
Consensus quality: 145204 bases at least Q30
Consensus quality: 145258 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 145367; sum-of-contigs
Quality coverage: 12.1 in Q20 bases; agarose-fp
Quality coverage: 12.1 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 75: contig of 75 bp in length
* 76 175: gap of 100 bp
* 176 13700: contig of 13525 bp in length
* 13701 13800: gap of 100 bp
* 13801 46229: contig of 32429 bp in length
* 46230 46329: gap of 100 bp
* 46330 145667: contig of 99338 bp in length.

```

```

FEATURES
source
1..145667
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="CTD-2357J19"
/clone_lib="CITD Human BAC"
misc_feature 1..75
/notes="assembly_fragment"
misc_feature 176..13700
/notes="assembly_fragment"
misc_feature 13801..46229
/notes="assembly_fragment"
misc_feature 46330..145667
/notes="assembly_fragment"
BASE COUNT 45215 a 27244 c 27442 g 45464 t 302 others
ORIGIN
alignment_scores:
Quality: 99.50 Length: 141
Ratio: 1.345 Gaps: 5

```


1

Tue Sep 17 07:27:30 2002

```

alignment_scores:
  Quality: 99.50      Length: 270
  Ratio: 0.783       Gaps: 14
  Percent Similarity: 47.037  Percent Identity: 18.889

alignment_block:
US-09-674-779-2 x AC016944/rev ..

Align seg 1/1 to reverse of: AC016944 from: 1 to: 158886

28 ProGlnValSerProLeuLysThrProSerValLeuLeuThrLysAspLy 44
||||:||||:||||: ||| ||||||||| |||
145016 CCCTCCATCAGCGCTCCAGCCCGCCATCTGCTCTGT..... 144975

44 sileGlyAspHisThrHisGluHisAspGluSerValSerHisVal. 60
|||:||||:||||: ||| ||||||||| |||
144974 .CTTCCTTACACACACACACGACGACACACACACACATGCTT 144926

61 .....GlyLeuGlnAlaHisPheGluThrTrpLeu 70
|||:||||:||||: ||| ||||||||| |||
144925 AGGAGAAACAGGAACACAGAGGATTCACACATTTATTTGCTTGG... 144879

71 GlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLe 87
||||:||||:||||: ||| ||||||||| |||
144878 CAATCTACCACTGACCTAAGAACCCAGGACCTTTGTTGCT.....TT 144835

87 uGlnSerArgLeuGly.....AsnTyrLeuProPromSer. 99
||||:||||:||||: ||| ||||||||| |||
144784 GAATCTCTTTTATTCTGACTAGATGATGATGATGATGATGATGAT 144735

112 Gly.....His..... 113
||||:||||:||||: ||| ||||||||| |||
144734 GGTGCTGCTGTAGGCGAGCTGTACATCTGGGAGACATTCATAGTGT 144585

114 GluProTyrGlnLeuProGluHisLeuTrpGlyGlnIle..... 127
|||:||||:||||: ||| ||||||||| |||
144684 GAGCACTGACTCTGTCTATCCCGACATAGAGGAGCAATGTAGTGGTG 144635

128 .....ValProThrLeuHisLeuTyrGln 135
||||:||||:||||: ||| ||||||||| |||
144634 CTCTGTGCCCCCTATGAGCAGCCCAATTCACAAATTCATGGGAGAA 144585

136 Asp.....LeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleAr 150
|||:||||:||||: ||| ||||||||| |||
144584 GGGACACGGCTGCTGAGCACACTGTGATCCACACACACTGAGTGCAG 144535

150 gSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyAlaAlaM 167
|||:||||:||||: ||| ||||||||| |||
144534 GGAT.....CCTGGGG 144524

167 etSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProaspLeu 183
|||:||||:||||: ||| ||||||||| |||
144523 TGAATAATGATGTTAATACATCACCCTGGCAGCATTTTCATCAGATGAA 144474

184 GluIleLysSer.....GlnAlaLeuTyrGluLeuGlnAsnAr 196
|||:||||:||||: ||| ||||||||| |||
144473 CAGATGGAAGCCCTGCAGCATCTCAGGGAGCGCTGGGAAATCAACACAA 144424

196 gLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuG 213
|||:||||:||||: ||| ||||||||| |||
144423 TGCTGCGTATATTTCTAGAT...GGGCAAAATAAAAC..... 144387

213 lLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLys 229
|||:||||:||||: ||| ||||||||| |||
144386 .....AAA 144384

230 TrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuPr 246
|||:||||:||||: ||| ||||||||| |||
144383 CCRAAAGAGGTTTAAAGGACGATGATGACGAGGAAGACGACATCAGCAC 144334

```

```

246 oLysAsnLys 249
|||||
144333 AATAAACAAA 144324

```

```

seq_name: gb_pr:AL139038

```

```

seq_documentation_block:

```

```

LOCUS      AL139038                140756 bp    DNA    linear    PRI 15-JUL-2001
DEFINITION Human DNA sequence from clone RP11-456B18 on chromosome 13,
complete sequence.

```

```

ACCESSION  AL139038
VERSION    AL139038.18  GI:14800148
KEYWORDS   HTG.
SOURCE     human.

```

```

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE  1 (bases 1 to 140756)
AUTHORS   Dunn,M.
TITLE     Direct Submission
JOURNAL   Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

```

```

requests:  clonerequest@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced gi:13751261.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

```

```

COMMENT

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13> RP11-456B18 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-456B18 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP11-456B18 is at 140756 in this sequence. The true left end of clone RP11-78L16 is at 68441 in this sequence. The true right end of clone RP11-108H9 is at 100 in this sequence.

FEATURES

source

```

1. .140756
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="13"
   /clone="RP11-456B18"
   /clone_lib="RPCI-11.2"
   1830..2326

```

repeat_region

```

/note="LORLa repeat: matches 1. .497 of consensus"

```

repeat_region

```

3932..3965
/note="17 copies 2 mer gt 94% conserved"

```

repeat_region

```

4632..5067
/note="LORLb repeat: matches 1. .461 of consensus"

```

repeat_region

```

5733..5893
/note="LIME3 repeat: matches 6002. .6146 of consensus"

```

```

repeat_region 6000. .6083
/Note="MER34 repeat: matches 454. .539 of consensus"
repeat_region 6108. .6171
/Note="HERV23 repeat: matches 993. .1056 of consensus"
repeat_region 7143. .7840
/Note="LTR8 repeat: matches 1. .691 of consensus"
repeat_region 8059. .8292
/Note="LTR16A repeat: matches 203. .442 of consensus"
repeat_region 8491. .8686
/Note="LTR29 repeat: matches 6. .203 of consensus"
repeat_region 8725. .8926
/Note="MER4D repeat: matches 483. .677 of consensus"
repeat_region 8992. .9049
/Note="MER4D repeat: matches 410. .465 of consensus"
repeat_region 9062. .9281
/Note="110 copies 2 mer tt 59% conserved"
repeat_region 9398. .9530
/Note="MER41C repeat: matches 1. .131 of consensus"
repeat_region 9623. .9732
/Note="LTR1 repeat: matches 675. .785 of consensus"
repeat_region 9970. .10333
/Note="LTR2D repeat: matches 1. .486 of consensus"
repeat_region 10400. .10545
/Note="LTR29 repeat: matches 2. .151 of consensus"
repeat_region 10549. .10935
/Note="MER67D repeat: matches 12. .391 of consensus"
repeat_region 11249. .11332
/Note="MIR repeat: matches 69. .144 of consensus"
repeat_region 11923. .12159
/Note="MER4D repeat: matches 1. .234 of consensus"
repeat_region 12164. .12767
/Note="MER4D repeat: matches 396. .973 of consensus"
repeat_region 12852. .13221
/Note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 13222. .14798
/Note="THE1B-INTERNAL repeat: matches 1. .1580 of consensus"
repeat_region 14799. .15173
/Note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 16144. .16346
/Note="MERS4B repeat: matches 585. .789 of consensus"
repeat_region 16339. .16503
/Note="MERS4B repeat: matches 112. .278 of consensus"
repeat_region 16504. .16571
/Note="MER66A repeat: matches 1. .69 of consensus"
repeat_region 16570. .16629
/Note="MER41B repeat: matches 283. .341 of consensus"
repeat_region 16599. .16902
/Note="MER66A repeat: matches 163. .478 of consensus"
repeat_region 16903. .17014
/Note="MERS4B repeat: matches 3. .115 of consensus"
repeat_region 17195. .17491
/Note="AluY repeat: matches 1. .297 of consensus"
repeat_region 18120. .19291
/Note="LIP repeat: matches 1726. .2897 of consensus"
repeat_region 19460. .19584
/Note="25 copies 5 mer atttt 62% conserved"
repeat_region 19686. .19745
/Note="12 copies 5 mer tttta 71% conserved"
repeat_region 20453. .20914
/Note="L1MD2 repeat: matches 5745. .6199 of consensus"
repeat_region 20913. .21222
/Note="L1MD2 repeat: matches 5277. .5594 of consensus"
repeat_region 21308. .21570
/Note="LTR9 repeat: matches 357. .625 of consensus"
repeat_region 21645. .21984
/Note="LTR19B repeat: matches 55. .393 of consensus"
repeat_region 22081. .22583
/Note="L1MD2 repeat: matches 4726. .5242 of consensus"
repeat_region 23498. .23784
/Note="AluB repeat: matches 1. .305 of consensus"
repeat_region 23873. .24169
/Note="AluSg repeat: matches 1. .295 of consensus"

repeat_region 25044. .25238
/Note="MIR repeat: matches 31. .238 of consensus"
repeat_region 25404. .25696
/Note="AluSx repeat: matches 20. .311 of consensus"
repeat_region 25832. .26218
/Note="MSTC repeat: matches 1. .405 of consensus"
repeat_region 26221. .27865
/Note="MSTC-internal repeat: matches 2. .1651 of consensus"
repeat_region 27907. .28659
/Note="L1PB2 repeat: matches 5400. .6155 of consensus"
repeat_region 28710. .29043
/Note="MSTD repeat: matches 30. .394 of consensus"
repeat_region 29634. .29697
/Note="32 copies 2 mer ta 79% conserved"
repeat_region 29729. .29752
/Note="12 copies 2 mer tg 95% conserved"
repeat_region 30688. .30874
/Note="MER61A repeat: matches 160. .354 of consensus"
repeat_region 31648. .31785
/Note="L1MC4 repeat: matches 6409. .6540 of consensus"
repeat_region 31813. .32243
/Note="L1MC4 repeat: matches 6629. .7089 of consensus"
repeat_region 32449. .32821
/Note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 32824. .34464
/Note="THE1C-internal repeat: matches 3. .1651 of consensus"
repeat_region 34465. .34825
/Note="THE1C repeat: matches 1. .360 of consensus"
repeat_region 34804. .35004
/Note="MER67C repeat: matches 202. .403 of consensus"
repeat_region 39792. .39868
/Note="L1PA12 repeat: matches 6076. .6152 of consensus"
repeat_region 40953. .41258
/Note="AluY repeat: matches 1. .308 of consensus"
repeat_region 41279. .41308
/Note="15 copies 2 mer aa 86% conserved"
repeat_region 42744. .42933
/Note="L1MA8 repeat: matches 6064. .6256 of consensus"
repeat_region 42973. .43275
/Note="AluSx repeat: matches 1. .303 of consensus"
misc_feature complement(43560. .43991)
/Note="match: STS: Em:HSPH07ES"
repeat_region 45196. .45584
/Note="MIR2B repeat: matches 2. .399 of consensus"
repeat_region 45594. .45767
/Note="87 copies 2 mer ta 77% conserved"
repeat_region 45800. .45851
/Note="MLT2B repeat: matches 394. .448 of consensus"
repeat_region 48118. .48487
/Note="THE1C repeat: matches 1. .370 of consensus"
repeat_region 48708. .48806
/Note="AluJo/FRAM repeat: matches 197. .297 of consensus"
repeat_region 49853. .50271
/Note="L1MEC repeat: matches 2274. .2369 of consensus"
repeat_region 50637. .50927
/Note="L1ME repeat: matches 974. .1253 of consensus"
repeat_region 52250. .52349
/Note="MIR repeat: matches 131. .228 of consensus"
repeat_region 52696. .52996
/Note="AluSg repeat: matches 1. .293 of consensus"
repeat_region 54927. .55010

alignment_scores:
Quality: 99.00 Length: 183
Ratio: 1.138 Gaps: 8
Percent Similarity: 47.541 Percent Identity: 25.683

alignment_block:
US-09-674-779-2 x AL139038/rev ..
Align seg 1/1 to reverse of: AL139038 from: 1 to: 140756

```



```

143154 AGCTTTCCAAAGTCTCTAAGAAGTTCCAAAGTTCCTACATATTTCTCCTGTGT 143105
      ::::: ::::: ::::: :::::
112 GlyHisGluProTyrGlnLeuProGluHisLeuTrpGlyGlnIleVa 128
      TCT: ::::: :::::
143104 TCTTCTGAGCCCTCCAAAGCTTTCCAAACCTCTGCCTTACTCAATTCCA 143055
      ::::: :::::
128 lProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProA 145
      ::::: :::::
143054 AAGCTGCTTCCACATTTTTCAGTATCTTTACAGCAGAGCCATACTA.... 143009
      ::::: :::::
145 laSnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCys 161
      TCT: :::::
143008 .....TCTCGTACCAATTTTACTGTATTA.....GTC 142982
      ::::: :::::
162 AlaGlyGlyAlaAlaMetSerLysHis..... 170
      ::::: :::::
142981 TGTTCATGCCGCTATGATAATAATACCTCAGACTGGGTAAATTAATAAAG 142932
      ::::: :::::
171 .....LeuThrAsnSerAlaIleAspIleTrpValPro 181
      ::::: :::::
142931 AAACAAAAGGTTTAAATGACTCAGCTTTTGACGGGCTGTGGGGCCT 142883
      ::::: :::::
seq_name: gb_hlg:AC103346

seq_documentation_block:
LOCUS AC103346 30062 bp DNA linear HTG 24-NOV-2001
DEFINITION Rattus norvegicus clone CH230-2K7, *** SEQUENCING IN PROGRESS ***,
9 unordered pieces.
AC103346
VERSION AC103346.1 GI:17063114
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 30062)
Muzny D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Kocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Hulys,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudas,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peary,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherier,S., Scott,G., Shen,H., Shooastari,N.,
Sisson,I., Suttergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vascuez,L., Vera,V., Villalon,D., Vinson,P.,

```

Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 30062)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Submitted (24-NOV-2001)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: TUV0

Center clone name: CH230-2K7

Summary Statistics

Sequencing vector: Plasmid; M77789

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 68432 bases at least Q40

Consensus quality: 72999 bases at least Q30

Consensus quality: 77137 bases at least Q20

Estimated insert size: 61291; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 0.7x in Q20 bases; agarose-fp estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 6200: contig of 6200 bp in length

* 6201 6300: gap of unknown length

* 6301 9304: contig of 3004 bp in length

* 9305 9404: gap of unknown length

* 9405 13486: contig of 4082 bp in length

* 13487 13586: gap of unknown length

* 13587 17093: contig of 3507 bp in length

* 17094 17193: gap of unknown length

* 17194 19456: contig of 2263 bp in length

* 19457 22347: contig of 2791 bp in length

* 22348 22447: gap of unknown length

* 22448 24839: contig of 2392 bp in length

* 24840 24940: gap of unknown length

* 24941 27217: contig of 2277 bp in length

* 27217 30062: contig of 2746 bp in length.

* Location/Qualifiers

* 1..30062

* /organism="Rattus norvegicus"

* /db_xref="taxon:10116"

* /clone="CH230-2K7"

BASE COUNT

7816 a 6888 c 6714 g 7802 t 842 others

ORIGIN

alignment_scores:

Quality: 98.50

Ratio: 1.145

Percent Similarity: 52.121

Percent Identity: 26.667

alignment_block:

US-09-674-779-2 x AC103346

Align seg 1/1 to: AC103346 from: 1 to: 30062

21 SerAlaProIleProThrAsnProGlnValSerProIleLeuThrProSe 37

14376 GCTGCTGTGTCGCCAGCAGTGACAAATACCAAGCAAGCTGTTCAGCCTTC 14425

37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54

14426 CTTCTAGTAGGAGCTGCTTTTACACACACACACACACACGACGCA 14475

54 spGluSerValSerHisValGlyLeuGlnAlaHisPheGlu..... 67

14476 CGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 14525

68ThrTrpLeuGlnMethHis.....HisAlaTh 76

14526 AGTAATAATGATGAGGCGGCACATGTTGAGAGTTGAAGTGCAATTGTTT 14575

76 rLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93

14576 CAGAGCTGGAGTTGAGATGG.....CAGAGCAGAAATAGTTT 14613

93 snTyrLeuProMetSerGlnLeuLeuThrThrAlaArgSerTrpGl 109

14614 CAGGAATAGATTGATGATGAG.....ACTTTCAAGCAGCAATGGCT 14657

109 nAlaCysGlyHisGluProTyrGlnLeuProGluHisLeuTrpGlyG 126

14658 TCTT.....GTGCTCCCTCCACCTGGCCATTGTTT.... 14688

126 InIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIle 142

14689CTGCTACAGGACCTGGAGTCCCGGACACCC 14718

143 LeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAs 159

14719 TTTCCTTCACACAGAGAACTAGGAGTAAACCTTCCAGGCAATTCCTAGA 14768

159 nGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsn 173

14769 T...ATCACCCTGCTGCTTGCATTAGTTCATCTGCATCA 14808

seq_name: gb_htg.AC013934

seq_documentation_block:

LOCUS AC013934 169516 bp DNA linear HTG 16-NOV-1999

DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***

ACCESSION AC013934

VERSION AC013934.1 GI:6437401

KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 169516)

Adams, M. and Venter, J.C.

Direct Submission

Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

This sequence was identified as CDM:10214156 by the submitter.

For further information on this sequence e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

Location/Qualifiers

1..169516

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

FEATURES

source

Align seg 1/1 to: AC009905 from: 1 to: 170662

17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProI1 33
|||||
15805 CTGTCGCATCGAGTCTCCG...GCTTGGCGGCTGAGAGTTCACCT... 165849
|||||
33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH1st 50
|||||
165850CGACGGGATCGTAATGCTGATGGGCATC 165877
|||||
50 hHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
|||||
165878 TTGCGGAG.....CGCATTCGACATGGCGGAGTGGAGCATATAT 165921
|||||
67 GluThrTriPLeuGlnMethHisAlaThrLysGlnGluVal.....Va 81
|||||
165922 CATCCATGGGTATGAGGACGGGATCTTGAAGCAATCTGTAAGCTTGTG 165971
|||||
81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProM 98
|||||
165972 TAGTGGCAG.....CGTCATCGAAGCTATCTCCACATC. 166005
|||||
98 etSerGlnLeuThrThrAlaArgSerTyrPbAlaCysGlyHisGlu 114
|||||
166006GCAGGCGGCACGAC 166020
|||||
115 ProTyrGlnLeuProProGluHis...LeuTyrGlyGlnIleVal..... 128
|||||
166021 CCT...GAAGTCCGCACTCATTTTGTATGGTATCTTGTACGTATCC 166067
|||||
129ProThrLeuHisLeuTyrG 135
|||||
166068 CTCTGAACCTTTGAAGTCCGTGTATCTATCTATCCAGTTCGGCCACGGTCTC 166117
|||||
135 lNAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
|||||
166118 AAGGAGTTGGTCAAGGTCCTTATCTGCTTGGATATCTCAAGGCGGAGT 166167
|||||
152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
|||||
166168 CCAAAACATCTGCTCAGACAGTGGGACTGGGCTCATGGACTTGTCTGTCA 166217
|||||
168 rLysHisLeuThrAsnSerAlaIle...AspIleTyrValProAspLeuGlu 184
|||||
166218 ATCTCACTGAGCTTCATCCACTCGGTGATCTTATCGTGGCGTCAAAAG 166267
|||||
185 lLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr... 200
|||||
166268 ACGTGGCTCAACTGCTACCAAAATCCAG.....TGTCTCAAGTTTC 166311
|||||
201TriPLeuGluHisGlyGluAlaGlnAsnPheGlyLeuGlyLeuT 215
|||||
166312 GTACCTTTTCATCGAATAGTGTCAATCTCTTTTCTGCTCGGATCGC 166361
|||||
215 yAlaThrGly 218
|||||
166362 AGCGGATGGGC 166372

seq_name: gb_in:AC010121

seq_documentation_block:
LOCUS AC010121 172529 bp DNA linear INV 21-APR-2001
DEFINITION Drosophila melanogaster, chromosome 2L, region 36C-36D, BAC clone
BACR07M13, complete sequence.
ACCESSION AC010121
VERSION AC010121.7 GI:13702787
KEYWORDS HTG
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.


```

* 34381 34480: gap of unknown length
* 34481 35515: contig of 1035 bp in length
* 35515 35615: gap of unknown length
* 35615 37145: contig of 1530 bp in length
* 37145 37245: gap of unknown length
* 37245 38635: contig of 1410 bp in length
* 38635 38755: gap of unknown length
* 38755 40418: contig of 1663 bp in length
* 40418 40518: gap of unknown length
* 40518 42089: contig of 1571 bp in length
* 42089 42189: gap of unknown length
* 42189 43485: contig of 1296 bp in length
* 43485 43585: gap of unknown length
* 43585 45537: contig of 1972 bp in length
* 45537 45558: gap of unknown length
* 45558 47561: contig of 1904 bp in length
* 47561 47661: gap of unknown length
* 47661 50270: contig of 2609 bp in length
* 50270 50370: gap of unknown length
* 50370 52463: contig of 2093 bp in length
* 52463 52563: gap of unknown length
* 52563 54241: contig of 1678 bp in length
* 54241 54341: gap of unknown length
* 54341 56282: contig of 1941 bp in length
* 56282 56382: gap of unknown length
* 56382 57724: contig of 1342 bp in length
* 57724 57824: gap of unknown length
* 57824 60404: contig of 2580 bp in length
* 60404 60504: gap of unknown length
* 60504 63040: contig of 2536 bp in length
* 63040 63140: gap of unknown length
* 63140 64315: contig of 1174 bp in length
* 64315 64415: gap of unknown length
* 64415 68063: contig of 3649 bp in length
* 68063 68163: gap of unknown length
* 68163 72715: contig of 4552 bp in length
* 72715 72815: gap of unknown length
* 72815 79858: contig of 7043 bp in length
* 79858 79958: gap of unknown length
* 79958 87305: contig of 7347 bp in length
* 87305 87405: gap of unknown length
* 87405 94304: contig of 6899 bp in length
* 94304 94404: gap of unknown length
* 94405 99973: contig of 5569 bp in length
* 99973 100073: gap of unknown length
* 100073 110741: contig of 10668 bp in length
* 110741 110841: gap of unknown length
* 110841 116556: contig of 5715 bp in length
* 116556 116656: gap of unknown length
* 116656 126137: contig of 9481 bp in length
* 126137 126237: gap of unknown length
* 126237 139661: contig of 13424 bp in length
* 139661 139761: gap of unknown length
* 139761 147923: contig of 8162 bp in length
* 147923 148023: gap of unknown length
* 148023 160030: contig of 12007 bp in length
* 160030 160130: gap of unknown length
* 160130 175441: contig of 15311 bp in length
* 175441 175541: gap of unknown length
* 175541 191133: contig of 15592 bp in length
* 191133 191233: gap of unknown length
* 191233 231912: contig of 40679 bp in length.

```

FEATURES

```

source
  1. .231912
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="RP23-127H3"
    /clone_lib="RP23 mouse BAC library 23"

```

```

BASE COUNT 61588 a 52985 c 59076 g 49983 t 8280 others
ORIGIN

```

alignment_scores:

```

Quality: 98.50      Length: 159
Ratio: 1.407        Gaps: 7
Percent Similarity: 44.025 Percent Identity: 27.673

alignment_block:
US-09-674-779-2 x AC087566 ..
Align seg 1/1 to: AC087566 from: 1 to: 231912

49 HistHrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHi 65
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196633 CATACGACACATGAT.....CACACAGATTGTGTCACACA 196670

65 sPheGluThrTrpLeuGlnMetHisAlaThrLysGlnGluValVala 82
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196671 CATGAATCATGCTGTCAGCTCTCCAGGCCACCTCCATCCCTGGTGA 196720

82 rGtyrGlnAlaTyrrLeuGlnSerArgLeuGlyAsnTyrrLeuProMet 98
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196721 TTCATCAGCACTTATGGGAATCTTCCCAAGACTGGGTCAATCCACAGCACC 196770

99 SerGlnLeu.....LeuThrThrAlaArgSerTrpGlnAlaCysGlyHi 113
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196771 AGCCACAGCCCTCCCTCTCCACTGCTGCTGCTGAAGGAATGTCAAGA 196820

113 sGluProTyrrGlnLeuProProGluHisLeuTrpGlyGlnLeuValProf 130
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196821 TGAGTTAATCCAG..... 196833

130 hrLeuHisLeuTyrrGlnAspLeuLysSerArgGlyLeuProAlaAsn 146
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196834 ..CCCCACATGAACAG.....AAGAAAT 196854

147 ThrGlnIleArgSerValTyrr.....ArgAsnProGluLe 158
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196855 AAGCAGACCAGAGAGGTAAAGTGAAGTGTTCGTGGCCGACAGCTAAITA 196904

158 uAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSera 175
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196905 GAACACAGATAAGTCCACCTCTGAAGCTTCT.....TCTG 196939

175 laIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrr 191
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196940 CCTCTCCGGAAGCACTCCAGGCTCACTGTGACATCTCAG..... 196980

192 GluLeuGlnAsnArgLeuCysGlnTyrr 200
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
196981 .....AGGCTCTGTGGCTAT 196995

seq_name: gb_in:AE003655

```

seq_documentation_block:

```

LOCUS AE003655 253753 bp DNA linear INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386055 section 48
of 63, complete sequence.
ACCESSION AE003655 AE002690
VERSION AE003655.2 GI:10728844
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 253753)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Fandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,

```

Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Folsler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpén, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Liu, X., Mattel, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleib, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P., Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weissenbach, G. M., Weissenbach, J., Williams, S. M., Woodage, T., Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Ye, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M. and Venter, J. C.

The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
2 (bases 1 to 253753)
Adams, M. D., Celniker, S. E., Gibbs, R. A., Rubin, G. M. and Venter, C. J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7298378.

CDS

```

1. 253753
   /organism="Drosophila melanogaster"
   /db_xref="taxon:7227"
   /chromosome="2L"
   join(327..428,1031..1082,1148..1482,10017..10079,
15983..16138,16199..16414,16480..16645,16700..16877,
17103..17168,17962..18072,20317..20427,20943..21009,
21072..22528,22583..24412,24480..24610,24678..24886,
24946..25286)
   /gene="CLIP-190"
   /product="CT16052"
   /db_xref="FLYBASE:FBan0005020"
   /db_xref="FLYBASE:FBgn0020503"
   <327..>25286
   /gene="CLIP-190"
   /note="CG5020"
   /map="36C1-36C2"
   /db_xref="FLYBASE:FBan0005020"
   /db_xref="FLYBASE:FBgn0020503"
   join(520..1082,1148..1482,10017..10079,15983..16138,
16199..16414,16480..16645,16700..16877,17103..17168,
17962..18072,20317..20427,20943..21009,21072..22528,
22583..24412,24480..24610,24678..24886,24946..25286)
   /gene="CLIP-190"
   /note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
   /product="CT15940"
   /db_xref="FLYBASE:FBan0005020"
   /db_xref="FLYBASE:FBgn0020503"
   join(1218..1482,10017..10079,15983..16138,16199..16414,
16480..16645,16700..16877,17103..17168,17962..18072,
20317..20427,20943..21009,21072..22528,22583..24412,
24480..24610,24678..24886,24946..24992)
   /gene="CLIP-190"

```

```

/note="CLIP-190 gene product [alt 1]: Nucleotide sequence
of the Celera sequence differs from the published sequence
for this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0005020"
/db_xref="FLYBASE:FBgn0020503"
/protein_id="AAF53605.1"
/db_xref="GI:7298379"
/translation="MSDDTSASGGTSAPFPSPVTADPPPGATASKLPGPIRSNIPTPA
TSCTGIPQPSKMKAPSGSTGVSIGRPGCCNHTTPKSGPPPREATSMRESDDNLS
SINSAYTNSSAVLTAQTEQFIIGQVLMGTGPGQIAFIGDTHFAAGWAGVILDEP
NGKNDGCVSGKRYFQCEPKRGIFSLRTLTITPLAGATPTSPILAKSPDPSRVSPT
ASIRSMRLSPGIGKNGMAGDVRIVISGFGSRPGILRYLGCTQFAPGNWCVGLDE
PSGKNDGTVDDIRYFCKPKYGVFPIAKVSLSPSKKTRLSRTSGRESLSICTMNS
IATYATSRMNAQAOKKSTPVKPIATPKPSQFMSQDILLREKQHQVEMKLVMDLRE
DAQNALQKNINELKARIVEALSALDNERRKTEELQKQVQVQVQVQVQVQVQVQV
KEKIHDLSEKTKLVSAATPSQSLIPDLPSDQALOEIAKLOEKMTIQOKEVESRI
AQLEFQRLRENRYLNEQATLSQELSVSKDEALEKFSLECGFNRLRELEILKEE
NRKQAEQAQAEFTRKLAKSVNRLSELQNLKATSDSLESERVNKTDECEILQTEV
RMRDQIRELNQLDEVTTQNVKADSSALDDMLRLQEGTEPEKSTLLEKTELQV
SKEOAKTLNDKEQLEKISDLKALQAEQKLVRETNAINQIQLEKESIQQLAKQ
NELEDFQKQSESEVHLQRIKAQNTKDFELVSGESLKKLQQLQLEKTELQQLAA
LEELKKEKETIIKEKEQLEQLQSKSAESALKVYVQVQVQVQVQVQVQVQVQVQV
AKLHDEISQKSAQAEQSELSKQLEAANAELKVNKEYAESRAESADLOKVEITK
LKSEVGETQAALSSCHTDVSEKTKQLEAANAELKVNKEYAESRAESADLOKVEITK
QLODSQDQTKLAKGEEKSEFESINLOVEVTKAKTENLELSTGTQTTIKLOER
LITNAELQHKEMASDAQIADLKLVEALQVANANISATNLELSTGTQTTIKLOER
ETNHIFELFEMEADMSRLTKVIGKEELKETHLOIDREKKEFEELEKQKQKQKQ
EQKLOQESQTEKLETKIQOQLQSDQVKEELVQNLQLEKRVESIIIEAQTINKL
ESNVOLNKTSLCTQDQLLESQKKEQLOEAAKLSGELQVQVQVQVQVQVQVQVQV
BELVKVLEKQAATSQLDAQAQATNKEQLELVKSOENEGNLOGESLAVTEKLOQLEQ
ANGELKEALCOKENGLKEQKLDSENTVLESOKSHNEIQDKLEQAQOKERTLOEET
SKLQQLQKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
LEQONQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
SLSGKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
DASGKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
DLTKRKPAPRLFCIDFQDQEDTCIQGSDQDQSTPSSSENNNEKPKLPAPR
KYDCSEVFGHDTSECADDEY"
   join(1218..1482,10017..10079,15983..16138,16199..16414,
16480..16645,16700..16877,17103..17168,17962..18072,
20317..20427,20943..21009,21072..22528,22583..24412,
24480..24610,24678..24886,24946..24992)
   /gene="CLIP-190"
   /note="CLIP-190 gene product [alt 2]"
   /codon_start=1
   /db_xref="FLYBASE:FBan0005020"
   /db_xref="FLYBASE:FBgn0020503"
   /protein_id="AAF53605.1"
   /db_xref="GI:7298379"
   /translation="MSDDTSASGGTSAPFPSPVTADPPPGATASKLPGPIRSNIPTPA
TSCTGIPQPSKMKAPSGSTGVSIGRPGCCNHTTPKSGPPPREATSMRESDDNLS
SINSAYTNSSAVLTAQTEQFIIGQVLMGTGPGQIAFIGDTHFAAGWAGVILDEP
NGKNDGCVSGKRYFQCEPKRGIFSLRTLTITPLAGATPTSPILAKSPDPSRVSPT
ASIRSMRLSPGIGKNGMAGDVRIVISGFGSRPGILRYLGCTQFAPGNWCVGLDE
PSGKNDGTVDDIRYFCKPKYGVFPIAKVSLSPSKKTRLSRTSGRESLSICTMNS
IATYATSRMNAQAOKKSTPVKPIATPKPSQFMSQDILLREKQHQVEMKLVMDLRE
DAQNALQKNINELKARIVEALSALDNERRKTEELQKQVQVQVQVQVQVQVQVQV
KEKIHDLSEKTKLVSAATPSQSLIPDLPSDQALOEIAKLOEKMTIQOKEVESRI
AQLEFQRLRENRYLNEQATLSQELSVSKDEALEKFSLECGFNRLRELEILKEE
NRKQAEQAQAEFTRKLAKSVNRLSELQNLKATSDSLESERVNKTDECEILQTEV
RMRDQIRELNQLDEVTTQNVKADSSALDDMLRLQEGTEPEKSTLLEKTELQV
SKEOAKTLNDKEQLEKISDLKALQAEQKLVRETNAINQIQLEKESIQQLAKQ
NELEDFQKQSESEVHLQRIKAQNTKDFELVSGESLKKLQQLQLEKTELQQLAA
LEELKKEKETIIKEKEQLEQLQSKSAESALKVYVQVQVQVQVQVQVQVQVQVQV
AKLHDEISQKSAQAEQSELSKQLEAANAELKVNKEYAESRAESADLOKVEITK
LKSEVGETQAALSSCHTDVSEKTKQLEAANAELKVNKEYAESRAESADLOKVEITK
QLODSQDQTKLAKGEEKSEFESINLOVEVTKAKTENLELSTGTQTTIKLOER
LITNAELQHKEMASDAQIADLKLVEALQVANANISATNLELSTGTQTTIKLOER
ETNHIFELFEMEADMSRLTKVIGKEELKETHLOIDREKKEFEELEKQKQKQKQ
EQKLOQESQTEKLETKIQOQLQSDQVKEELVQNLQLEKRVESIIIEAQTINKL
ESNVOLNKTSLCTQDQLLESQKKEQLOEAAKLSGELQVQVQVQVQVQVQVQVQV
BELVKVLEKQAATSQLDAQAQATNKEQLELVKSOENEGNLOGESLAVTEKLOQLEQ
ANGELKEALCOKENGLKEQKLDSENTVLESOKSHNEIQDKLEQAQOKERTLOEET
SKLQQLQKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
LEQONQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
SLSGKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
DASGKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
DLTKRKPAPRLFCIDFQDQEDTCIQGSDQDQSTPSSSENNNEKPKLPAPR
KYDCSEVFGHDTSECADDEY"
   join(1218..1482,10017..10079,15983..16138,16199..16414,
16480..16645,16700..16877,17103..17168,17962..18072,
20317..20427,20943..21009,21072..22528,22583..24412,
24480..24610,24678..24886,24946..24992)
   /gene="CLIP-190"

```

CDS


```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.1-26"
/clone="RP1-278N12"
/clone_lib="RPC1-1"
1. .121
repeat_region
/notes="L1M4 repeat: matches 5913. .6034 of consensus"
122. .412
repeat_region
/notes="AluSg repeat: matches 1. .290 of consensus"
413. .669
repeat_region
/notes="L1M4 repeat: matches 6034. .6286 of consensus"
706. .853
repeat_region
/notes="AluSg/X repeat: matches 135. .299 of consensus"
857. .1167
repeat_region
/notes="AluSx repeat: matches 1. .308 of consensus"
1501. .1812
repeat_region
/notes="AluJo repeat: matches 3. .310 of consensus"
2652. .2949
repeat_region
/notes="AluSx repeat: matches 1. .300 of consensus"
2972. .3082
repeat_region
/notes="L2 repeat: matches 2572. .2697 of consensus"
3620. .3924
repeat_region
/notes="AluSg repeat: matches 1. .273 of consensus"
4005. .4175
repeat_region
/notes="TIGER1 repeat: matches 2238. .2416 of consensus"
4176. .4659
repeat_region
/notes="L1M4 repeat: matches 5784. .6285 of consensus"
4685. .5034
repeat_region
/notes="L1M3 repeat: matches 4910. .5235 of consensus"
5035. .5342
repeat_region
/notes="AluSg repeat: matches 1. .307 of consensus"
5343. .6094
repeat_region
/notes="L1M3 repeat: matches 5235. .6139 of consensus"
6335. .6508
repeat_region
/notes="AluJb repeat: matches 134. .311 of consensus"
8493. .8614
repeat_region
/notes="FLAM_A repeat: matches 1. .122 of consensus"
8616. .8691
repeat_region
/notes="38 copies 2 mer ta 71% conserved"
8700. .8997
repeat_region
/notes="AluJo repeat: matches 1. .300 of consensus"
9263. .9573
repeat_region
/notes="AluSx repeat: matches 1. .309 of consensus"
9584. .10001
repeat_region
/notes="L1M4 repeat: matches 3585. .4012 of consensus"
10003. .10136
repeat_region
/notes="AluJb repeat: matches 2. .135 of consensus"
10090. .10460
misc_feature
/notes="match: STS Z23751; genomic marker D6S425"
10139. .10168
repeat_region
/notes="15 copies 2 mer ac 97% conserved"
10170. .10307
repeat_region
/notes="L1M4 repeat: matches 4063. .4203 of consensus"
10323. .10616
repeat_region
/notes="AluSg repeat: matches 1. .296 of consensus"
10625. .10812
repeat_region
/notes="L1M3 repeat: matches 7481. .7675 of consensus"
complement(10716. .11105)
misc_feature
/notes="match: GSS AQ218681"
10931. .11376
repeat_region
/notes="LTR16C repeat: matches 1. .387 of consensus"
11377. .11439
repeat_region
/notes="L1M3 repeat: matches 7718. .7779 of consensus"
11440. .11725
repeat_region
/notes="AluSg repeat: matches 3. .288 of consensus"
11726. .11855
repeat_region
/notes="L1M3 repeat: matches 7779. .7913 of consensus"
12052. .12192
repeat_region
/notes="FLAM_C repeat: matches 2. .138 of consensus"
12293. .12590
repeat_region
/notes="AluSx repeat: matches 1. .298 of consensus"
12608. .12943
/notes="L1M4 repeat: matches 5125. .5467 of consensus"
12820. .13220
misc_feature
/notes="match: GSS B80534"
13039. .13120
repeat_region
/notes="41 copies 2 mer ta 73% conserved"
13312. .13449
repeat_region
/notes="L2 repeat: matches 2612. .2745 of consensus"
13711. .14015
repeat_region
/notes="AluSg repeat: matches 1. .307 of consensus"
14053. .15990
repeat_region
/notes="L1P1A1 repeat: matches 4207. .6144 of consensus"
16225. .16366
repeat_region
/notes="FLAM_C repeat: matches 1. .133 of consensus"
16898. .17018
repeat_region
/notes="FLAM_C repeat: matches 2. .122 of consensus"
17081. .17383
repeat_region
/notes="AluJo repeat: matches 1. .308 of consensus"
17384. .17421
repeat_region
/notes="19 copies 2 mer ga 82% conserved"
17669. .17925
misc_feature
/notes="match: GSS AQ527764"
18093. .18388
repeat_region
/notes="AluSx repeat: matches 1. .309 of consensus"
18669. .18790
repeat_region
/notes="MER50 repeat: matches 1. .129 of consensus"
18793. .19028
repeat_region
/notes="AluSx repeat: matches 42. .276 of consensus"
19030. .19271
repeat_region
/notes="121 copies 2 mer aa 65% conserved"
19309. .19898
repeat_region
/notes="MER50 repeat: matches 117. .711 of consensus"
complement(20062. .20712)
misc_feature
/notes="match: ESTs AI269581 AA226135 AI476542 AA226134"
20064. .20162
repeat_region
/notes="FLAM_C repeat: matches 1. .133 of consensus"
20399. .20867
misc_feature
/notes="match: GSS AQ499824"
20934. .21330
repeat_region
/notes="MLT2B repeat: matches 1. .404 of consensus"
21333. .21462
repeat_region
/notes="65 copies 2 mer at 76% conserved"
21464. .21599
repeat_region
/notes="68 copies 2 mer at 76% conserved"
22069. .22142
repeat_region
/notes="MIR repeat: matches 68. .143 of consensus"
22451. .22530
repeat_region
/notes="MIR repeat: matches 52. .131 of consensus"
22583. .23228
repeat_region
/notes="LTR29 repeat: matches 5. .619 of consensus"
23254. .23320
repeat_region
/notes="L1M4A repeat: matches 5149. .5211 of consensus"
23321. .23448
repeat_region
/notes="FLAM_C repeat: matches 1. .127 of consensus"
23449. .23462
repeat_region
/notes="L1M4A repeat: matches 5211. .5224 of consensus"
23461. .24364
repeat_region
/notes="L1M4A repeat: matches 5191. .6045 of consensus"
24365. .24670
repeat_region
/notes="AluJb repeat: matches 1. .305 of consensus"
24671. .24917
repeat_region
/notes="L1M4A repeat: matches 6045. .6300 of consensus"
complement(25321. .25698)
misc_feature
/notes="match: EST H75280"
complement(25359. .25909)
misc_feature
/notes="match: GSS AQ592624"
25533. .25673
repeat_region
/notes="LTR23 repeat: matches 1. .146 of consensus"
25717. .25796
repeat_region
/notes="MIR repeat: matches 52. .131 of consensus"
26104. .26177
repeat_region
/notes="MIR repeat: matches 68. .143 of consensus"
26431. .26726
repeat_region
/notes="AluJb repeat: matches 27. .304 of consensus"
```

```

misc_feature      complement(26669..27116)
                  /note="match: GSS A0138816"
repeat_region     26727..27005
                  /note="LTR23 repeat: matches 135..489 of consensus"
repeat_region     27207..27338
                  /note="FLAM_C repeat: matches 2..133 of consensus"
repeat_region     27347..27663
                  /note="AluJ0 repeat: matches 1..312 of consensus"
repeat_region     27855..28043
                  /note="MER4B repeat: matches 481..610 of consensus"
repeat_region     28004..28318

```

```

alignment_scores:
  Quality: 98.00      Length: 271
  Ratio: 0.803       Gaps: 15
Percent Similarity: 45.018 Percent Identity: 22.509

```

```

alignment_block:
US-09-674-779-2 x HS278N12

```

```

Align seg 1/1 to: HS278N12 from: 1 to: 78533

```

```

22 AlaProIleProThrAsnProGlnValSerProIleLysThrProSerVa 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42082 TCACCTCTACCATCACTTCTCCACCATCTCCACCATGTCTACCATCACT 42131
38 lleu.....lleThrLysAspLysIleGlyAspHisHisThrHisGluH 53
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42132 TCTACCATCTGTCACTCCACCGTCACCTCCACACCATCACTCCACCA 42181
53 lsAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrp 69
|||:|||||:|||||:|||||:|||||:|||||:|||||:
42182 ACCTCCACCAACCATCTCCACCATCTCTCCACCAACCATCTCCACCAT 42231
70 leu.....GlnMetHisHisAla..... 75
|||||:|||||:|||||:|||||:|||||:|||||:
42232 CTCACACACCATCTCCACCATCTCTACCGTCACCTCCACCATCACTCTA 42281
76 ....ThrLysGlnGluValAlaArgTyrGlnAlaTyrLeuGlnSerArgL 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42282 CCACCTCCAATCAACCATTTCTACCATCTCCATTCACCATCTCCACCAT 42331
91 euGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThrAlaArgSer 107
|||:|||||:|||||:|||||:|||||:|||||:|||||:
42332 TCACCAACCACTCTACCATCTCCACCATCTCCACCATCTCCACCAT 42360
108 TrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGlu..... 121
|||||:|||||:|||||:|||||:|||||:|||||:
42361 .....CATTCCACCATCACTCCCTCCACCATCACTT 42395
122 .....HisLeuTrpGlyGlnIleValP 129
|||||:|||||:|||||:|||||:|||||:|||||:
42396 CCAACACCAACCCCTTTATCTCCACCTCCACCTCCACCATGTATCGC 42445
129 roThrLeuHisLeuTyr.....GlnAspLeuLysSerArg 140
|||||:|||||:|||||:|||||:|||||:|||||:
42446 CAACCTTTCACTTTACTGACCTGGGGATTTTCTGTATCCCGCTTCACA 42495
141 Gly.....IleLeuProAlaAsnThrG1 148
|||||:|||||:|||||:|||||:|||||:|||||:
42496 CACCTTTCTTACTCCACACTTCATATCTCTCTCTCTCTCTCTCTCTCT 42545
148 nileArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGly. 164
|||||:|||||:|||||:|||||:|||||:|||||:
42546 GATC.....TTTTCTAATCACTTCACTTTTAAAGTCCTGAGTAAACA 42586
165 .....AlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
|||||:|||||:|||||:|||||:|||||:|||||:
42587 GAGATAAAGCAATGGCT.....TTC 42606
180 ValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnAr 196
|||||:|||||:|||||:|||||:|||||:|||||:
42607 CTACCA.....CATTCATTTTATGAA..... 42627

```

TITLE
JOURNAL

```

196 gluCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuG 213
|||||:|||||:|||||:|||||:|||||:|||||:
42628 .TTATGCCAGTTATTTCATT...CATTCACAAAT..... 42657
213 lylLeuTyrAlaThrGlyAlaIleHisLeuAspThrClnGlyPheArgLys 229
|||||:|||||:|||||:|||||:|||||:|||||:
42658 .....ATTACTGAGACGACGACGACGACGACGACGACGACGACGAC 42687
230 TrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuPr 246
|||||:|||||:|||||:|||||:|||||:|||||:
42688 CTGGGAGC.....ACAACTACTCTCTCATTTGCAGGATTCCTTTAA 42731
246 oLysAsnLysLeu 250
|||||:|||||:|||||:|||||:|||||:
42732 GGCTAAAGAGCTT 42744

```

```

seq_name: gb_htg:AC097543

```

```

seq_documentation_block:
LOCUS      AC097543
DEFINITION Rattus norvegicus clone CH230-13611, *** SEQUENCING IN PROGRESS
*** 74 unordered pieces.
ACCESSION  AC097543
VERSION    AC097543.4 GI:17973388
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

```

```

REFERENCE  1 (bases 1 to 146321)

```

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Briefe,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Cartoll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegod,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nickerson,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 146321)

Worley, K.C.

Direct Submission

Submitted (19-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17064342.

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GHUV

Center clone name: CH230-12611

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 118791 bases at least Q40

Consensus quality: 127201 bases at least Q30

Consensus quality: 132476 bases at least Q20

Estimated insert size: 94009; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 74 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 5167: contig of 5167 bp in length

* 5168 5267: gap of unknown length

* 5268 8729: contig of 3462 bp in length

* 8730 8829: gap of unknown length

* 8830 12812: contig of 3983 bp in length

* 12813 12912: gap of unknown length

* 12913 16451: contig of 3539 bp in length

* 16452 16551: gap of unknown length

* 16552 19971: contig of 3420 bp in length

* 19972 20071: gap of unknown length

* 20072 23215: contig of 3144 bp in length

* 23216 23315: gap of unknown length

* 23316 27807: contig of 4492 bp in length

* 27808 27907: gap of unknown length

* 27908 30577: contig of 2670 bp in length

* 30578 30677: gap of unknown length

* 30678 34125: contig of 3448 bp in length

* 34126 34225: gap of unknown length

* 34226 36057: contig of 1832 bp in length

* 36058 36157: gap of unknown length

* 36158 38083: contig of 1926 bp in length

* 38084 38183: gap of unknown length

* 38184 40595: contig of 2412 bp in length

* 40596 40695: gap of unknown length

* 40696 41732: contig of 1037 bp in length

* 41733 41832: gap of unknown length

* 41833 44795: contig of 2963 bp in length

* 44796 44895: gap of unknown length

* 44896 47736: contig of 2841 bp in length

* 47737 47836: gap of unknown length

* 47837 49547: contig of 1711 bp in length

* 49548 49647: gap of unknown length

* 49648 51800: contig of 2153 bp in length

* 51801 51900: gap of unknown length

* 51901 54832: contig of 2932 bp in length

* 54833 54932: gap of unknown length

* 54933 57421: contig of 2489 bp in length

* 57422 57521: gap of unknown length

* 57522 59541: contig of 2120 bp in length

* 59542 59741: gap of unknown length

* 59742 61775: contig of 2034 bp in length

* 61776 61875: gap of unknown length

* 61876 63221: contig of 1346 bp in length

* 63222 63321: gap of unknown length

* 63322 64796: contig of 1475 bp in length

* 64797 64896: gap of unknown length

* 64897 66314: contig of 1418 bp in length

* 66315 66414: gap of unknown length

* 66415 68369: contig of 1955 bp in length

* 68370 68469: gap of unknown length

* 68470 70263: contig of 1794 bp in length

* 70264 70363: gap of unknown length

* 70364 72846: contig of 2483 bp in length

* 72847 72946: gap of unknown length

* 72947 75430: contig of 2484 bp in length

* 75431 75530: gap of unknown length

* 75531 77680: contig of 2150 bp in length

* 77681 79155: contig of 1375 bp in length

* 79156 79255: gap of unknown length

* 79256 81496: gap of unknown length

* 81496 83077: contig of 1482 bp in length

* 83078 83177: gap of unknown length

* 83178 84399: contig of 1222 bp in length

* 84400 84500: gap of unknown length

* 84500 85780: contig of 1281 bp in length

* 85781 85880: gap of unknown length

* 85881 87000: contig of 1120 bp in length

* 87001 87101: gap of unknown length

* 87101 88556: contig of 1556 bp in length

* 88557 88756: gap of unknown length

* 88757 89448: contig of 1092 bp in length

* 89449 89949: contig of 2092 bp in length

* 89949 92041: gap of unknown length

* 92041 92141: contig of 1879 bp in length

* 92141 94120: gap of unknown length

* 94120 95784: contig of 1665 bp in length

* 95785 95884: gap of unknown length

* 95885 97071: contig of 1187 bp in length

* 97072 97171: gap of unknown length

* 97171 98387: contig of 1216 bp in length

* 98388 98488: gap of unknown length

* 98488 99735: contig of 1247 bp in length

* 99735 99834: gap of unknown length

* 99835 101634: contig of 1800 bp in length

* 101635 101734: gap of unknown length

* 101735 102788: contig of 1053 bp in length

* 102788 102888: gap of unknown length

* 102888 104248: contig of 1360 bp in length

* 104248 104347: gap of unknown length

* 104348 105668: contig of 1321 bp in length

* 105669 105768: gap of unknown length

* 105769 107032: contig of 1264 bp in length

* 107033 107132: gap of unknown length

* 107133 108500: contig of 1368 bp in length

* 108501 108600: gap of unknown length

* 108601 110253: contig of 1653 bp in length

* 110254 110353: gap of unknown length

* 110354 112309: contig of 1956 bp in length

* 112310 112409: gap of unknown length

* 112410 113605: contig of 1196 bp in length

* 113606 113705: gap of unknown length

* 113706 114730: contig of 1025 bp in length

* 114731 114830: gap of unknown length

* 114831 116310: contig of 1480 bp in length

* 116311 116410: gap of unknown length

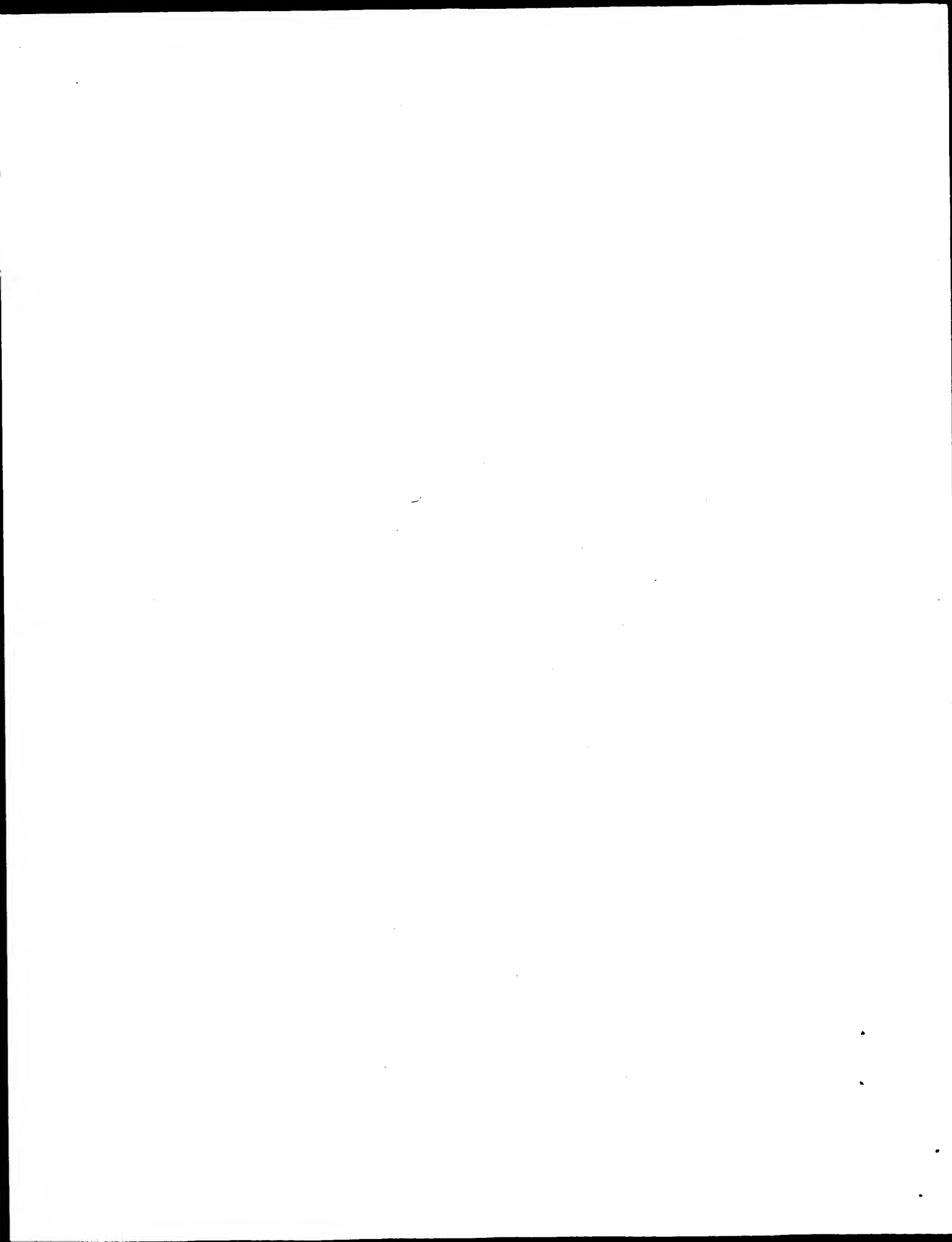
* 116411 117540: contig of 1130 bp in length

* 117541 117641: gap of unknown length

* 117641 118994: contig of 1354 bp in length

901 TTGCATTTCAGCTATGAACACGRCATATGCTAAAGAGAGCTGCGACTA 950
185 eLysSerGlnAla.....LeuTyrGluLeuGlnAsnArgL 197
951 TCACAGCCCGCAGTCGCCCTTATCGTTATGAAGGAGCTAGAGAATCTAT 1000
197 euCysGlnTyrTrpLeuGluHis 204
1001 CGAATAAAAGCCTGTGTTGACCAT 1023

CC	* be preserved.
CC	* 1 1066: contig of 1066 bp in length.
XX	
XX	
FH	Key Location/Qualifiers
FH	
FT	source
FT	1.._L066
FT	/db_xref="taxon:5741"
FT	/organism="Giardia intestinalis"
FT	/strain="WB-C6"
FT	/clone="AJ1492"
XX	
SQ	Sequence 1066 BP; 312 A; 270 C; 224 G; 259 T; 1 other;
alignment_scores:	
	Quality: 97.50 Length: 241
	Ratio: 0.871 Gaps: 12
	Percent Similarity: 46.473 Percent Identity: 22.822
alignment_block:	
US-09-674-779-2 x AC034780 ..	
Align seg 1/1 to: AC034780 from: 1 to: 1066	
4	PheAsnGlnTyrPheIleThrLeuIleSerMetLeuValAlaCys 20
376	TATAATGAGTTGTCAAGACGAAGTACAGGATAACCCGGTTGTTCTCTG 425
20	s.....SerAlaProIleProThrAsnProGlnValSerProIleL 34
426	TCAGTGCTCTTTAGTAGTGATTGCACAGCCTTCTCAGAGCTCCAAGCGGT 475
34	yThrProSerValLeuIleThrLysAspLysIleGlyAspHisThr 50
476	CAACCCCTTCGTACGAGATTACATCTGGTAGT...GGAGATTATTCTGTA 522
51	HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheCl 67
523	..GAGAGAAACATATGCTAACAACACCTGAAC..... 552
67	uThrTrpLeuGlnMethHisAlaThrLysGlnGlu..... 79
553CATATGCACGAATGCTAGCCCTGGCTCGGCC 586
80ValValArgTyrGlnAlaTyrLeuGlnSerArgLeu..... 91
587	TTTCTCTCATCGTAGGACACAAGAAGCAATCCAGCTTAAGCTTAGTGCC 636
92GlyAsnTyrLeuProProMetSerGlnLeuLeuThrAla 105
637	ATGAAAACATCGCATACCACCGAGATATGCTGTTCTCTG..... 678
105	aArgSerTrpGlnAlacysGlyHisGluProTyrGlnLeuProProGluHu 122
679CACGATTCCCTAGGTATTAAAGACTATTC 706
122	IsLeuTrp.....GlyGlnIleValProThrLeuHis 132
707	ATCCTTGGGTTTCGGGGGCACCTCTTCTCTCTGTTATTTCAACAACCCCTGCAC 756
133	Leu.....TyrGlnAspLeuLysSerArgGlyIleLeuProAlas 146
757	TTGGGAGATTGTCAATAGAGAATCTGAGATAGGCATCGCTCCCTCAA 806
146	nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlag 163
807	TACAGACAGGCAA.....ATGTCAAACTCTCAAGTTGATGACTGTGCGC 850
163	Iy.....GlyAlaAlaMetSer 168
851	ATACCCAGCGCTCCGTGAATTCACCTCTCAGACAGTCCACATAGCATGGCC 900
169	IvsHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIl 185



OM of: US-09-674-779-2 to: N_Geneseq_032802.* out_format : pfs
Date: Sep 16, 2002 6:57 PM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framer-p2n.model -DEV=xlh
-Q/cn2_1/USPTO_spo01/US09674779/runat_12092002.124122.25875/app_query.fasta_1.310
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=std.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LLOPCL=0.000 -LLOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTENT=pfs
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09674779 -CGENL_1.165 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-2
Query length: 250
Database: N_Geneseq_032802.*
Database sequences: 1756436
Database length: 858457221
Search time (sec): 199.900000

score.list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59797	..	1355.00	2647.89	3.8e-1	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF28530	..	1355.00	2609.26	5.5e-1	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABL20561	..	98.50	152.39	3.85	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL20561	..	98.50	139.23	20.85	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL11490	..	98.50	138.85	20.85	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL11502	..	98.50	138.60	22.59	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS36567	..	95.00	153.50	3.34	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI59893	..	94.00	158.76	1.70	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAL05039	..	93.50	127.50	93.81	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS45146	..	92.00	150.56	4.87	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV34374	..	92.00	128.95	77.88	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH6027	..	89.50	147.11	7.58	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL15895	..	89.50	133.16	35.10	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL15779	..	89.50	129.73	42.85	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABL15779	..	89.50	136.41	70.47	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS35758	..	89.50	118.46	299.15	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS35754	..	89.50	118.42	300.68	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS35759	..	89.50	112.94	607.25	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS35759	..	89.50	112.94	607.25	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAD10354	..	89.50	140.54	17.61	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAS02149	..	88.50	146.30	8.41	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS03562	..	88.50	144.63	10.42	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAL06616	..	88.50	144.63	10.42	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAH81475	..	88.00	149.99	5.24	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH81475	..	88.00	149.22	5.78	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAH84500	..	88.00	149.22	5.78	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAS03352	..	88.00	149.06	5.90	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL03309	..	88.00	148.64	6.23	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS46271	..	88.00	120.46	231.35	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA82625	..	88.00	101.43	2.7e-0	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK80624	..	87.00	142.96	12.90	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK80625	..	87.00	126.91	101.15	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL25734	..	87.00	126.89	101.35	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV18082	..	86.00	130.61	62.91	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH81475	..	86.00	141.18	16.22	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT:AAV77394	..	86.00	139.46	20.22	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT:AAH81083	..	86.00	138.89	21.75	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAS75095	..	86.00	137.06	27.52	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:ABL32230	..	86.00	114.67	485.81	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:AAT11658	..	86.00	109.79	909.37	

/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:ABA05882 - 86.00 109.79 909.37
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI99683 - 86.00 50.32 1.4
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI99682 - 86.00 50.30 1.4
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ53602 + 85.50 144.87 10.1
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:AAQ46887 + 85.50 135.05 35.6

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59797

seq_documentation_block:

ID AAF59797 standard; DNA; 753 BP.

XX AAF59797;

XX 04-MAY-2001 (first entry)

XX Moraxella catarrhalis strain ATCC43617 BASB120 DNA.

XX BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
XX genetic immunisation; infection; upper respiratory tract; otitis media;
XX hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
XX invasive disease; antibacterial; auditory; ds.

XX Moraxella catarrhalis.

XX WO200109335-A2.

XX 08-FEB-2001.

XX 31-JUL-2000; 2000WO-EP07361.

XX 03-AUG-1999; 99CA-0018281.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Thonnard J;

WPI; 2001-159872/16.

P-PSDB; AAB60645.

New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis strain American Type Culture Collection 43617, for use as therapeutic agents or vaccines against bacterial infections, e.g. otitis media or pneumonia

Claim 13; Page 64; 75pp; English.

The invention relates to the Moraxella catarrhalis strain ATCC43617 BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The invention also relates to immunogenic fragments of the BASB120 protein, expression vectors and host cells comprising BASB120 nucleic acids, the recombinant production of BASB120, vaccine compositions comprising the BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic compositions comprising the anti-BASB120 antibody, and a method of identifying a Moraxella catarrhalis infection via the detection of BASB120 proteins or antibodies. The vaccine compositions of the invention are useful as prophylactic or therapeutic agents against Moraxella catarrhalis infections in mammals, particularly humans. Moraxella catarrhalis is a Gram negative bacterium frequently isolated from the human upper respiratory tract, which is responsible for several pathological conditions. It is responsible for about 15% of otitis media cases in children (which can lead to temporary or permanent hearing loss). It also causes pneumonia in elderly people, and sinusitis, nosocomial infections and, less frequently, invasive diseases. BASB120 proteins or nucleotides may additionally be used in screening for novel antibacterial compounds, and in the diagnosis and staging of infections. The present sequence represents DNA encoding the Moraxella catarrhalis strain ATCC43617 BASB120 protein.

Sequence 753 BP; 239 A; 166 C; 152 G; 196 T; 0 other;

alignment_scores:

Quality: 1355.00

Length: 250

Ratio: 5.420 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x AAF59797 ..
Align seg 1/1 to: AAF59797 from: 1 to: 753
1 MetLysAsnPheAsnGlnTyrPheIleThrLeuIleSerSerMetLe 17
|||||
1 ATGAAAAATTTTAAATCAATCTTTATAACTATACACTTATCAGCAGTATGCT 50
17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
|||||
51 GGTTCATGTCAGTCACCCATACCAACCAATCCACAGTATCCCAATTA 100
34 yThrProSerValLeuIleThrLysAspLysIleGlyAspHisHsThr 50
|||||
101 AACGCCATCGGTACTGATTACTTAAGATATAAATCGGTGATCATCATACA 150
51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
|||||
151 CATGAGCACGATGAATCTGTAAGCCATGTCGGTTGCGAGGCACATTTGA 200
67 uThrTrpLeuGlnMethHisAlaThrLysGlnGluValValArgTyrG 84
|||||
201 GACTTGGCTACAGATGACCATGCCACCAACAGAGGTAGTTAGTATC 250
84 lAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
|||||
251 AGGCGTATTATCAATCAAGACTTGGTAATATCTGCCACCAATGAGTCAA 300
101 LeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG1 117
|||||
301 CTACTAAGCAGTGCACCGTATGCGAGGCGATGGTGCATGAACCTTATCA 350
117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeu 134
|||||
351 GCTGCCACCAACAATCTTTGGGGTCAGATTGTACCAACATGTCACATTGT 400
134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
|||||
401 ATCAAGATCTCAAAAGTAGGGGCATATGCGCAGCAATACCAAAATTCGC 450
151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
|||||
451 TCAGTTTATCAGCAATCTCGAATCAACCAATGCTGCTGGTGCAGCTAT 500
167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
|||||
501 GAGTAAGCATTTGACCAATAGTGCATTTGATTTGGTGGCTGACCTTG 550
184 lIleLeuSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
|||||
551 AAATAAAGCCAGGCACTGTATGAGCTTCAAAACCGCCTATGCCAATAT 600
201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
|||||
601 TGGCTAGAGCATGGCGAAACCAAAATTTGGGCTGGGTTTATACGCCAC 650
217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
|||||
651 AGGGCGGATTCATCTGGATACCCCAAGGGTTTAGAAAAATGGGGTCTCAAT 700
234 hSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
|||||
701 TTTCTGAAACAACACTATTGTCGTCTATGCTTACCAAAAAATAAGCTA 750
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF28530
seq_documentation_block:
ID AAF28530 standard; DNA: 23210 BP.
XX

AC AAF28530;
XX 04-APR-2001 (first entry)
DT
XX Genomic fragment #17.
DE
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX Moraxella catarrhalis.
OS
XX WO200078968-A2.
PN
XX 28-DEC-2000.
PD
XX 16-JUN-2000; 2000WO-US16649.
PF
XX 18-JUN-1999; 99US-0140121.
PR
XX (INCY-) INCYTE GENOMICS INC.
XX PA
XX Lagace RE, Patterson C, Berg KL;
PI
XX WPI: 2001-041427/05.
DR
XX Genomic library for identifying diagnostic and therapeutic
XX compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids
XX
XX Claim 1: Page 141-146; 545pp; English.
PS
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specifications e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 23210 BP; 7108 A; 4482 C; 4951 G; 6668 T; 1 other;

alignment_scores:
Quality: 1355.00 Length: 250
Ratio: 5.420 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-674-779-2 x AAF28530 ..
Align seg 1/1 to: AAF28530 from: 1 to: 23210
1 MetLysAsnPheAsnGlnTyrPheIleThrLeuIleSerSerMetLe 17
|||||
11483 ATGAAAAATTTTAAATCAATCTTTATAACTATACACTTATCAGCAGTATGCT 11532
17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
|||||
11533 GGTTCATGTCAGTCACCCATACCAACCAATCCACAGTATCCCAATAA 11582
34 yThrProSerValLeuIleThrLysAspLysIleGlyAspHisHsThr 50
|||||
11583 AAACGCCATCGGTACTGATTACTTAAGATATAAATCGGTGATCATCATACA 11632
51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
|||||
11633 CATGAGCAGCATGAATCTGTAGCCATGTCGGTTGCGAGGCACATTTGA 11682
67 uThrTrpLeuGlnMethHisAlaThrLysGlnGluValValArgTyrG 84

```

11583 GACTTGGCTACAGTACCATGCCACCAAGAGGAGTAGTGTATC 11732
84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGln 100
11733 AGCGTATTACAAATCAAGACTTGTAATATCTGCCCAATGAGTCAA 11782
101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
11783 CTACTAACGACTGCACGGTATCGGCAGGCATGTGTATGAACCTTATCA 11832
117 nLeuProProGluHisLeuTyrGlnLeuValProThrLeuHisLeuT 134
11833 GCTGCCACCAAGACATCTTTGGGTGAGATTGTACCAATTTGCACTTGT 11882
134 YrGlnAspLeuLysSerArgGlyLeuProAlaAsnThrGlnLeuArg 150
11883 ATCAAGATCTCAAAAGTAGGGCATATTGCCAGCAAAATACCCAAATTCGC 11932
151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
11933 TCAGTTTATCGCAATCCTGAATCAACCAATGTCTGTGTGTCAGCTAT 11982
167 tSerYsHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
11983 GAGTAAGCATTTGACCAATAGTGCATTCATATTTGGGTGCTGACCTTG 12032
184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
12033 AAATAAAAGCCAGGCACCTGTATGAGCTTCAAAACCGCTATGCCAATAT 12082
201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
12083 TGGTAGACATGGCGAAACCAAAATTTGGGCTGGGTTTATAGCCAC 12132
217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
12133 AGGGGCGATTATCTGTGATACCAAGGTTTAGAAATGGGGTCTCAAT 12182
234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
12183 TTTCGAAACAAACCTATTTGTCTGCTATGCTTACCAAAAAAATAAGCTA 12232
seq_name: /SIBS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL20561
seq_documentation_block:
ID ABL20561 standard; DNA; 4560 BP.
AC ABL20561;
AT
CT
GT
TT
26-MAR-2002 (first entry)
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 13156.
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
OS Drosophila melanogaster.
XX WO200171042-A2.
PN
PD
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR

```

```

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 13156; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 4560 BP; 1344 A; 1186 C; 1089 G; 941 T; 0 other;
SQ

```

alignment_scores:

```

Quality: 98.50 Length: 221
Ratio: 0.895 Gaps: 12
Percent Similarity: 49.774 Percent Identity: 25.792

```

alignment_block:

```
US-09-674-779-2 x ABL20561/rev ..
```

```
Align seq 1/1 to reverse of: ABL20561 from: 1 to: 4560
```

```

17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProI 33
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2455 CTGTCGCATCGGAGTCTCG...GCTGGGGCGCTGAGAGTTCACCT.. 2411
33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisHist 50
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2410 .....CGACGGGATCGTAATGCTGATGGGCATC 2383
50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2382 TTCGGGAG.....CGCATTCGCACATGGCGGAGTTGGAGGCATATAT 2339
67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGlnVal.....Va 81
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2338 CATCCATGGGTATGAGGACGGGATTCTTGAGGCAATCTGTAGCTTGTG 2289
81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPROM 98
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2288 TAGGTGGCAG.....CGTCATCGAAGCTATCTCCACCT.. 2255
98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2254 .....GCAGCGGCGCACGAC 2240
115 ProTyrGlnLeuProGluHis...LeuTrpGlyGlnIleVal..... 128
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2239 CCT...GAAGTCCCGCAACTCATTTTGTATGTTCTTGTACGTATCC 2193
129 .....ProThrLeuHisLeuTyrG 135
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2192 CTCTGAACCTTTGAAGTCGTTATATCTCATCCAGTTCGGCCACGCTCTC 2143
135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2142 AAGGATTCGTCAAGGGTCTTATCTGCTGATATCTCAAGGCGGAGT 2093
152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2092 CCAAAACATCTGCCTCAGCAGGTGGGACTGGGCTCATGGACTGCTGTTC 2043
168 rLysHisLeuThrAsnSerAlaIle..AspIleTrpValProAspLeuGlu 184

```


XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR WPI; 2001-656860/75.
 XX DR P-PSDB; ABB67387.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 XX PS Claim 1; SEQ ID NO 28952; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 15166 BP; 4466 A; 2981 C; 3067 G; 4652 T; 0 other;

alignment_scores:

Quality: 98.50 Length: 221
 Ratio: 0.895 Gaps: 12
 Percent Similarity: 49.774 Percent Identity: 25.792

alignment_block:

US-09-674-779-2 x ABL11490 ..
 Align seg 1/1 to: ABL11490 from: 1 to: 15166
 17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProI1 33
 4916 CTGGTCGCATCGAGCTCCG...GCTTGGGGGCTGAGAGTTCCACCT.. 4960
 33 elysThrProSerValLeulleThrLysAspLysileGlyAspHisHist 50
 4961CGACGGGATCGTAATGCTGATGGGCATC 4988
 50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
 4989 TTCGGGAG.....CGCATTCGCACATGGCGGAGTTGGAGGCATTATAT 5032
 67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal.....Va 81
 5033 CATCATGGGTATGAGGACGGGATCTTGAAGCAATCTGTAAGCTGTG 5082
 81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProm 98
 5083 TAGGTGGCAG.....CGTCATCGAAGCTATCTCCACCT.. 5116
 98 etSerGlnLeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
 5117GCAGGGGGCAGCAG 5131

115 ProTyrGlnLeuProGluHis....LeuTrpGlyGlnIleVal..... 128
 5132 CCT...GAAGTCCGCCCACTCATTTTGTATGATATTCTTTGTACGATCC 5178
 129ProThrLeuHisLeuTyrG 135
 5179 CTCTGAACCTTTGAAGTCCGTGTATATCTCATCCAGTTCCGCCACGGCTCTC 5228
 135 lAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
 5229 AAGGAGTTTCGTCGAAGGTCTTATCTGCCTGGATATCTCAAGGGCGAGT 5278
 152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
 5279 CCAAAACATCTGCTCAGCAGTGGGACTGGCTCATGGACTTGTCTGTCA 5328
 168 rLysHisLeuThrAsnSerAlaIle..AspIleTrpValProAspLeuGlu 184
 5329 ATCTCACTGGACTTCATCCACGCTGATCTTATCTGTCGCTGCGTCAAAGG 5378
 185 lLysSerGlnAlaLeuTyrGlnLeuGlnAsnArgLeuCysGlnTyr... 200
 5379 ACCTGGCTCCAACTGCTGACCCAAATCCAG.....TTGCTCCAGTTCTC 5422
 201TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu 215
 5423 GTACCTTTTCATCGATATGTTGTCATCTCTTTTCTGCTGCGATCGC 5472
 215 yrAlaThrGly 218
 5473 AGCGGATGGGC 5483

seq_name: /SID1/gcgdata/hold-geneseg/geneseq-emb1/NA2001B.DAT:ABL11502

seq_documentation_block:

ID ABL11502 standard; cDNA; 15504 BP.
 XX AC ABL11502;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 28988.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical; gene; ss.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US09231.
 XX PR 23-MAR-2000; 2000US-191637P.
 XX PR 11-JUL-2000; 2000US-0614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR WPI; 2001-656860/75.
 XX DR P-PSDB; ABB67399.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX PT interactions -
 XX PS Claim 1; SEQ ID NO 28988; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (AB101840-ABL16175) and the encoded proteins
 CC (AB16177-AB162072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 15504 BP; 4552 A; 3065 C; 3141 G; 4746 T; 0 other;

alignment_scores:
 Quality: 98.50 Length: 221
 Ratio: 0.895 Gaps: 12
 Percent Similarity: 49.774 Percent Identity: 25.792

alignment_block:
 US-09-674-779-2 x ABL11502 ..
 Align seg 1/1 to: ABL11502 from: 1 to: 15504

17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProI1 33
 |||||.....
 4916 CTGGTCGATCGGAGTCTCG...GCTTGGCGCTGAGAGTTCACCT... 4960

33 eLysThrProSerValLeuLeuThrLysAspLysIleGlyAspHisH1st 50
CGACGGGATCGTAATGCTGATGGGCATC 4988

50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66

 4989 TTCGGGAG.....CGCATTCGACATCGGGCGGAGTGGAGGCATTATAT 5032

67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal.....Va 81

 5033 CATCCATGGTATGAGGAGGATCTTGAAGGCAATCTGTAAGCTGTGTG 5082

81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPROM 98
CGTCATCGAAGCTATCTCCACCT... 5116

5083 TAGGTGGGAG.....
 98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
CGACGGCGGCACGAC 5131

5117
 115 ProTyrGlnLeuProProGluHis...LeuTrpGlyGlnIleVal..... 128
 |||.....
 5132 CCT...GAAGTCCCGCCCACTATTGTAAGTATGTTCTGATGATCC 5178

129ProThrLeuHisLeuTyrG 135

 5179 CTTCTGAACCTTGAAGTCCGTGATATCTATCATCCAGTTCGGCCACGGTCTC 5228

135 lAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
 |||.....
 5229 AAGGAGTTCGTCAAGGTCTTATCTGCTGATATCTCAAGGGCGGAGT 5278

152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168

 5279 CCAAAACATCTGCTCAGCAGTGGGCTGATGCTATGCTGCTGCTCA 5328

168 rLysHisLeuThrAsnSerAlaIle...AspIleTrpValProAspLeuGlu 184

 5329 ATCTCACTGACTTCATCCATCGGTGATCTTATCGTTGGCGCGTCAAG 5378

185 lLysSerGlnAlaLeuTyrGlnLeuGlnAsnArgLeuCysGlnTyr... 200
 |||.....
 5379 ACGTGGCTCAAGTCTGACCAATCCAG.....TTGCTCCAGTCTTC 5422

201TrpLeuGluHisGlyGlnAsnGlnAsnPheGlyLeuGlyLeuT 215

 5423 GTACCTTTTCATCGATAATGGTCAATCTCTTTTCTCTCTCGGATCGC 5472

5423 GTACCTTTTCATCGATAATGGTCAATCTCTTTTCTCTCTCGGATCGC 5472

215 yrAlaThrGly 218
 ||| |||
 5473 AGCGCATGGCGC 5483

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT: AAS36567

seq_documentation_block:
 ID AAS36567 standard; DNA: 2240 BP.
 XX
 AC AAS36567;
 XX
 DT 17-DEC-2001 (first entry)
 XX
 DE Human cardiovascular system antigen genomic DNA SEQ ID No 2067.
 DE
 XX
 KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
 KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
 KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
 KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
 KW hyperproliferative disorder; breast; liver; cardiovascular infection;
 KW cerebrovascular disorder; nervous system disorder; bacterial infection;
 KW fungal infection; viral infection; ocular disorder; endocrine disorder;
 KW gastrointestinal disorder; renal disorder; respiratory disorder;
 KW wound healing; skin aging; organ transplantation; tissue regeneration;
 KW anti-infertility.
 XX
 OS Homo sapiens.
 XX
 PN WO200155321-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01340.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.

```
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-451930/48.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
XX Claim 1; SEQ ID No 2067; 674pp; English.
XX
XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
alignment_scores:
Quality: 95.00 Length: 303
Ratio: 0.772 Gaps: 13
Percent Similarity: 40.594 Percent Identity: 19.802
alignment_block:
US-09-674-779-2 x AAS36567 ..
Align seg 1/1 to: AAS36567 from: 1 to: 2240
```


alignment_scores:

Quality: 94.00 Length: 200
 Ratio: 1.000 Gaps: 12
 Percent Similarity: 47.000 Percent Identity: 27.500

alignment_block:

US-09-674-779-2 x AAI59893/rev ..

Align seg 1/1 to reverse of: AAI59893 from: 1 to: 1179

```

75 AlathrlsGlnGluValValArgTyrGlnAlaTyrLeuGln..... 88
||||| :|||: :|||: :|||: :|||:
1131 GCAACCAACAATCAACCTACAGGTGCAAGGACAACCTCCAGGAGTCTC 1082
89 .....SerArgLeuGlyAsnTyrLeuProPromets 99
||||| :|||: :|||: :|||: :|||:
1081 AAGTGCAGCTGTGCTCAGGCAGATTTGGCATACACCCACCTACC 1032
99 erGlnLeu.....LeuThrThr 104
||| :|||: :|||: :|||: :|||:
1031 CTGAGGCACACAGGGGGAAGGAGGAGTGGGGGAGAGAGCTAAC 982
105 AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGl 121
:||||: :||||: :||||: :||||: :||||:
981 AGCAGAAAGGGCAGGACCGGCGC...CTTCCCAACAAACCTCCCAAC 935
121 uHisLeuTrpGly.GlnIle.....ValProThrLeuHisLeuTyr 134
||| :|||: :|||: :|||: :|||:
934 CTGGTGCCTGTGGTCAGCTGAGGTCAGGAAGCCCACTCACCACTGGTA 885
135 GlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
||| :|||: :|||: :|||: :|||:
884 CAGCGCTCAATCAATCAATGCCAGTCTGG...GCCTGCACAGCGGATGC 838
151 rValTyrArgAsnProGluLeuAsnGlnCys.....AlaGlyGlyAla 166
||| :|||: :|||: :|||: :|||:
837 GGTA.....CGCACAGCACTCGGGTCAGCTGGGGGTCGA 800
166 laMetSerLysHis.....LeuThrAsnSerAlaIleAsp 177
||| :|||: :|||: :|||: :|||:
799 GCCATCCAGGAGGAGAGACATTCACCCCAACAGCACTGCCTCT 750
178 IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuG 194
:||||: :||||: :||||: :||||: :||||:
749 TCTTTCCTGCCAGCAAAACCTGCAAGAGAGATCA.....GGATACA 706
194 naSnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhe 211
||||| :|||: :|||: :|||: :|||:
705 AAACAAAACATACATTAC.....CTGG 683
211 lyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe 227
||||| :|||: :|||: :|||: :|||:
682 GTCTCCCACTCAGGCAAGCTGCCACCATCAGAT...GAAGCAGC 636
228 ArgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
||||| :|||: :|||: :|||: :|||:
635 CGGAAGTGGGAAGTCCACACACAGTGTCTCTATCTGCAAAACAT 588

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI05039

seq_documentation_block:

ID AAL05039 standard; DNA; 17324 BP.

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

```

XX WO200155320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216847.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.
XX 26-SEP-2000; 2000US-0235484.
XX 27-SEP-2000; 2000US-0235834.

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI05039

seq_documentation_block:

ID AAL05039 standard; DNA; 17324 BP.

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

XX AC AAL05039;

```

27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
08-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251889.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX Disclosure; SEQ ID NO 7727; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX Sequence 17324 BP; 3826 A; 5147 C; 4719 G; 3632 T; 0 other;
SQ
alignment_scores:
Quality: 93.50 Length: 226
Ratio: 0.926 Gaps: 11
Percent Similarity: 44.690 Percent Identity: 23.451
alignment_block:
US-09-674-779-2 x AAL05039/rev ..
Align seg 1/1 to reverse of: AAL05039 from: 1 to: 17324
58 SerHisValGlyLeuGlnAlaHisPheGluThrTriLeuGlnMethHis1 74
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
7074 ACCCACCAGGGGATCCAGTCTTGGTGGAGAACTGCTGTCAGAAC... 7027
74 sAlaThrLysGlnGluValValArgTyr..... 83
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
7026 ....ACACCCGAGGAGATGCCCTTCTGTACAGGGCGAGGGCTGA 6981
84 .....GlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr... 94
6980 ACAAGACAGCCATCGGGAGTACTCTGGGGAGAGGTAGCTGTCACCTCA 6931
95 .....LeuProMetSerGlnLeuLeuThrAlaAr 106
6930 GCTCTGTGGGGCCCTCCCTCCACCCAGACAGCCAGCTCTCTGCCCTC 6881
106 gSerTrpGlnAlaCysGly.HisGluProTyrGlnLeuProGluHis 122
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6880 ACACCTGGCAGCTCACAGTGGAAACAGCCCTGCAGCCTCCCGCAGAGGG 6831
123 LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSe 139
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6830 CCCTGG.....CAGATCAGCCCTCTGACTGGACTGTGTACA 6796
139 rArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 156
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6795 GAATGGACTG.....ACATGGGCTAGGGAGCCTCGCGCCCAAC 6758
156 roGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
||:|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6757 CGAGAGCATCTGGGATGGGGCCCTTCTGGAAATTACCTGGCAGTTGG 6708
173 AsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAl 189
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6707 GGAGGC...ATCCATTATGTCTGAGAGGGTAGAACAATCTCTCAAGG 6661
189 aLeuTyrGluLeuGlnAsnArg.....LeuCysGlnTyrT 201
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6660 ATCATGGGAGCTGAGACCCGCTGGGTAGACCATGCTGTGTAGTACGACT 6611
201 rLeuGluHisGlyGlu..... 206
||:|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
6610 GGTGGCCATCTGGGAGCTGTGGGACCCCTGGGCCCTTGTGGGAGCTTA 6561
207 .....AsnG1 208
XX
```

```

6560 GGAGTTGTGGGGTCATTGGAAGCCCTAGGTCACAGGACTGGACATCA 6511
      :::||
208 nAsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaLeHisLeuAspThrG 225
      |::||| |::|::|::|::|
6510 GATGATTGGAGGCCATCTGTGGAGCTCTGGG.....AATC 6476
      :::|
225 InGlyPheArgLys...TrpGlyAla 232
      |||||:: ::|||::|
6475 AGGGATTTATAGAGATCTGGGGTGCT 6450
      |||||:: ::|||::|

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.AAS45146
seq_documentation_block:
ID AAS45146 standard; cDNA; 1721 BP.

```

18-DEC-2001 (first entry)

cdna encoding novel human secretory protein, Seq ID No 415.

Human: secreted protein; arthritis; Crohn's disease; sepsis; shock; ischaemia-reperfusion injury; haemopoiesis; cancer; neuropathy; transgenic animal; Alzheimer's disease; Parkinson's disease; burn; atrophic lateral sclerosis; platelet disorder; thrombocytopenia; ulcer; osteoporosis; bone degenerative disorder; periodontal disease; gut protection; lung; liver fibrosis; immune deficiency; infection; severe combined immunodeficiency; SCID; autoimmune disorder; allergy; multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma; fertility; analgesic; pain; antigen; ss.

Homo sapiens.

WO200166689-A2

13-SEP-2007

	8
O	C
O	C
A	C
y	D
i	F
A	X
	.
e	E
I	C

03 MAR 2001, 2001W0-US0494Z.

U7-MAR-2000; 2000US-0519705.
19-MAY-2000; 2000US-0574454.

17-JUN-2000; 2000US-0596193.
14-JUL-2000; 2000US-0616847

19-SEP-2000; 2000US-0665363.

. /07CC00 000007 /0007 700-7

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;

WPI; 2001-589934/66.

04207044 / 0000

novel polypeptides and nucleic acids obtained from cDNA libraries prepared from various human tissues, for diagnosis and treatment of cancer, neurological, inflammatory, and autoimmune disorders -

Claim 1; SEQ ID No 415; 107pp; English.

The invention relates to novel isolated human secreted polypeptides (I) and polynucleotides (II). (I) and (II) are useful for treating inflammatory conditions such as arthritis, nephritis, Crohn's disease, ischaemia-reperfusion injury, shock, sepsis, immune responses, and is involved in increasing haematopoiesis, stem cell survival, bone growth and remodeling. (I), (II) and modulators of (II) are useful for prophylaxis or treatment of one or more cancers. (II) is also useful for creating transgenic animals useful for studying the *in vivo* activities of the polypeptide as well as for studying modulators of the polypeptides. (I) induces the proliferation of neural cells and regeneration of nerve and brain tissue and is useful for the treatment of central and peripheral nervous system diseases and neuropathies, such as Alzheimer's.

426 TCAC 423

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV33474

seq_documentation_block:

ID AAV33474 standard; cDNA; 11717 BP.

XX AC AAV33474;

XX DT 10-DEC-1998 (first entry)

XX DE Girdwood S.A.virus cDNA sequence.

XX KW Bone marrow cell; alphavirus; Girdwood S.A.virus; cytokine; capsid;
 reverse transcriptase-PCR; nsP1; nsP2; nsP3; nsP4; E3; E2; 6K; E1;
 growth hormone; growth factor; interleukin; chemokine; enzyme;
 ribozyme; antisense oligonucleotide; ss.

XX OS Girdwood S.A. virus.

XX FH Key Location/Qualifiers

XX FT 5'UTR 1..59

XX FT /*tag= a

XX FT CDS 60..7616

XX FT /*tag= b

XX FT /product= "Non-structural polyprotein"

XX FT /note= "Internal stop codon present at nucleotides

XX FT 5763-5765"

XX FT CDS 7662..11399

XX FT /*tag= c

XX FT /product= "Structural polyprotein"

XX FT 11400..11717

XX FT /*tag= d

XX PN WO9836779-A2.

XX PD 27-AUG-1998.

XX PF 18-FEB-1998; 98WO-US02945.

XX PR 19-FEB-1997; 97US-0801263.

XX PA (UYNC-) UNIV NORTH CAROLINA.

XX PI Davis NL, Johnston RE, Simpson DA;

XX XX WPI; 1998-495361/42.

XX DR P-PSDB; AAW70462, AAW70463, AAW70473, AAW70474, AAW70475, AAW70476,

XX DR AAW70477; P-PSDB; AAW70478, AAW70479.

XX PT Expression of heterologous RNA in bone marrow cells - using a
 recombinant alphavirus comprising a promoter operable in bone marrow
 cells operably associated with heterologous RNA.

XX PS Example 2; Fig 3A-3C; 68pp; English.

XX CC The invention provides a method for introducing and expressing
 heterologous RNA in bone marrow cells using alphavirus vectors. The
 Girdwood S.A.virus is an alphavirus whose cDNA sequence was determined
 from uncloned reverse transcriptase-PCR reaction fragments amplified
 from the virion RNA. The Girdwood S.A.virus cDNA encodes nonstructural
 and structural polyproteins. The nonstructural polyprotein is,
 presumably, post-translationally modified into four different protein
 products, namely nsP1 (AAW70462), nsP2 (AAW70463), nsP3 (AAW70473), nsP4
 (AAW70474) proteins. The structural polyprotein is, presumably, post-
 translationally modified into five different protein products, namely
 capsid (AAW70475), E3 (AAW70476), E2 (AAW70477), 6K (AAW70478), E1
 (AAW70479) proteins. The Girdwood S.A.virus cDNA was used in the method
 of the invention. The inventors claim the transformed bone marrow cells
 are useful for expressing a protein or peptide suitable for protecting
 the subject against a disease such as a microbial, bacterial, protozoal,
 parasitic or viral disease. The transformed bone marrow cells are also
 claimed to be useful for expressing proteins and peptides such as

CC hormones, growth hormones, growth factors, interleukins, cytokines,
 CC chemokines, enzymes, ribozymes or antisense oligonucleotides.

XX Sequence 11717 BP; 3297 A; 3130 C; 2927 G; 2360 T; 3 other;
 SQ

alignment_scores:
 Quality: 92.00 Length: 283
 Ratio: 0.724 Gaps: 18

Percent Similarity: 44.876 Percent Identity: 24.735

alignment_block:

US-09-674-779-2 x AAV33474 ..

Align seg 1/1 to: AAV33474 from: 1 to: 11717

11 ThrLeuLeuSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27

8664 ACCTTGACGAGCCGCTACTGGGCACATGCTGCTACTGTCCACCTACTGCA 8713
 ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||

27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44

8714 ACCGTGCTTTAGCCCGATTAAAGATCGAGCAGGTCTGG.....GATG 8754
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

44 ysileGlyAsp..HisHisThrHisGluHisaspGluSerValSerHisV 60

8755 AAGCGGACGACAAACACATACGATACAGACTTCCGCCAGTT..... 8797
 :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

60 aGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaThr 76

8798TGGATACGACCAAAAGCGGAGCAGC 8821
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

77 LysGlnGluValVal..ArgTyrGlnAlaTyrLeuGln..... 88

8822 AAGCTCAATAAGTACGCTACATGCTCGAGCAGGATCATACCGTCA 8871
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

88

8872 AAGAGGCATATGGATGACATCAAGATCAGCACTCAGGACCGCTGAGA 8921

89SerArgLeuGlyAsnTyrLeu.....ProPro..... 97

8922 AGGCTTAGCTACAAAGGATGCTTTCTCTCGCGAAGTGTCTCCAGGGGA 8971
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

98MetSerGlnLeuThrAlaArgSerTrpGlnAlaCys..... 111

8972 CAGGTAAACGGTTAGTAGCGAGTACCAACTCAGCAACGTCATGCACAA 9021
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

112

9022 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGACCTA 9071
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrG 135

9072 CTTCCCGTTAC.....GGTAAAGAGATCTCTTGCACA...GTGTACGA 9112
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

135 nAspLeuLys.....SerArgGlyIleLeu.....ProAla. 145

9113 CGGTCTGAAGAAACCAACCGCGCTCATCTACATGCACAGGCGGGAC 9162
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162

9163 GCGACGCTATACGCTCTCTATCTGGGAAATCATCAG..... 9198
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179

9199 .GGAAGTCTACGGCAACCCCAT..... 9222
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

179 tpValProAspLeuGluLeuLysSerGlnAla..LeuTyrGluLeuGlnAs 195

9223CCGAAAGAACATTAGCTAGTGAAGTGGCGGCGATTACAAGA 9267
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212


```

1199  GCCTCAGCGGCTCAACCTGGGAGTGCCTTTCTGCAAAAGCTGCTGCAC 1248
196  ArgLeuCysGlnTyrTrpLeuGluHis...GlyCyluAsnGlnAsnPheG1 211
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1249  GAC.....CTGGAACACAGTGGCGAGCGCCGCGAGTTCGA 1283
211  yLeuGlyLeuTyrAlaThr.....GlyAlaIleHisLeuAspThrG 225
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1284  TGCAGCGCTTGAGCGGTGCGCTCCCTGGCTGGCTGCTCCCTGGAGTCGC 1333
225  InGlyPhe 227
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1334  CCGGCTAC 1341
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL15895
seq_documentation_block:
ID ABL15895 standard; cDNA; 4360 BP.
XX
XX ABL15895;
XX
26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 42167.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmacetical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) FE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB7192.
XX

```

new isolated nucleic acid sequences for elucidating cell signalling and cell-cell interactions -

Claim 1; SEQ ID NO 42167; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB185737-AB187202).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 4360 BP; 1134 A; 1166 C; 1129 G; 931 T; 0 other;

alignment_scores:			
Quality:	89.50	Length:	168
Ratio:	1.065	Gaps:	8
Percent Similarity:	50.000	Percent Identity:	24.405

PR	17-NOV-2000;	2000US-02492008
PR	17-NOV-2000;	2000US-02492009
PR	17-NOV-2000;	2000US-02492010
PR	17-NOV-2000;	2000US-02492111
PR	17-NOV-2000;	2000US-02492112
PR	17-NOV-2000;	2000US-02492113
PR	17-NOV-2000;	2000US-02492114
PR	17-NOV-2000;	2000US-02492116
PR	17-NOV-2000;	2000US-02492117
PR	17-NOV-2000;	2000US-02492118
PR	17-NOV-2000;	2000US-02492244
PR	17-NOV-2000;	2000US-02492245
PR	17-NOV-2000;	2000US-02492264
PR	17-NOV-2000;	2000US-02492265
PR	17-NOV-2000;	2000US-02492297
PR	17-NOV-2000;	2000US-02492299
PR	01-DEC-2000;	2000US-02493000
PR	01-DEC-2000;	2000US-02501600
PR	01-DEC-2000;	2000US-02503391
PR	05-DEC-2000;	2000US-02510300
PR	05-DEC-2000;	2000US-02511988
PR	06-DEC-2000;	2000US-02567119
PR	06-DEC-2000;	2000US-02511479
PR	08-DEC-2000;	2000US-02511856
PR	08-DEC-2000;	2000US-02511868
PR	08-DEC-2000;	2000US-02511869
PR	08-DEC-2000;	2000US-02511899
PR	08-DEC-2000;	2000US-02515980
PR	11-DEC-2000;	2000US-02515990
PR	05-JAN-2001;	2001US-02540978
PR	05-JAN-2001;	2001US-02596788

Rosen CA, Barash SC, Ruben SM:

New cardiovascular system related polynucleotides and polypeptides, useful for diagnosing, treating and/or preventing disorders of the cardiovascular system -

Claim 1; SEQ ID No 1258; 674pp; English.

Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode the cardiovascular system antigen polypeptides of the invention

Cardiovascular system antigens and their associated polynucleotides are useful in the diagnosis, treatment and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by detecting the presence or absence of a mutation in a cardiovascular system antigen polynucleotide. The treatable disorders in cardiovascular diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma and pleurisy. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemoradiotherapy.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

alignment_scores:

Quality: 89.50

Ratio: 1.041

Length: 183

Gaps: 10

Percent Similarity: 46.995 Percent Identity: 25.683

```
alignment_block:
US-09-674-779-2 x AAS35758
..
Align seg 1/1 to: AAS35758 from: 1 to: 19206

10 ThrThrLeuLeuSerSerMetLeuValAlaCysSerAla..... 22
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
8970 ACCTGTTTGTCTCAGCTGGGGCTCTTAGCATGCTCTCTTCCTGGCTGGGTG 8919
23 .....ProtleProThra 27
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
8920 TGGGCTCTGCTTTCTCTGCGCCGCAAGCCCTGGGGTCCGGGTACCCCTGTT 8969
27 snProGlnValSerProIleLysThrProSerValLeuLeuThr..... 41
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
8970 CCTTTTCCACCATGCCAGGTGATCTCATCTCATCTGCATCAC.....CT 9057
42 ..LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVa 57
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9020 CCAGCCAGCCGCTCTCTGATCCATCATCTGCATCAC.....CT 9057
57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHis 74
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9058 TTCACATGTCCTTCAGCGGCCCTCAGACT..... 9091
74 isAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArg 90
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9092 .....CAAGCT.....GTCCAGATAGA 9109
91 LeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrThraArgSe 107
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9110 GTTGGC....CATGCCGCCACCATTTACCCAGATGTCACACCTGAAAGTCG 9156
107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGluHisLeuT 124
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9157 AGTTCTTGTCTCCACCATCATCTCATCTGCATCCCA..... 9193
124 rpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9194 .....TCGGTCTCTAGTCTGCTGCCCTAGTGCAA.....CTGCAA 9229
141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProG 157
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9230 TCCTTCTTGCCTCTGCTCTCTCTG..... 9253
157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
9254 .....CAGTGCCTAGTCTAGGCCACCATCTCTCAGTGCATCTCACA 9295
seq_name: /SIDSL/gcdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AAS35754
seq_documentation_block:
ID_ AAS35754 standard; DNA: 19274 BP.
XX
AC AAS35754;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 1254.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnery; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
OS Homo sapiens.
```

PR	27-SEP-2000;	2000US-02358336.
PR	29-SEP-2000;	2000US-02363367.
PR	29-SEP-2000;	2000US-02363367.
PR	29-SEP-2000;	2000US-02363368.
PR	29-SEP-2000;	2000US-02363369.
PR	29-SEP-2000;	2000US-02363370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0238602.
PR	20-OCT-2000;	2000US-0237038.
PR	20-OCT-2000;	2000US-0237039.
PR	20-OCT-2000;	2000US-0237040.
PR	20-OCT-2000;	2000US-0239935.
PR	20-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0241805.
PR	20-OCT-2000;	2000US-0241806.
PR	01-NOV-2000;	2000US-0246177.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246533.
PR	08-NOV-2000;	2000US-0246539.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251038.
PR	05-DEC-2000;	2000US-0251098.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2001US-0253497.
XX		
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		

```
1157 AGTCTTGTCCACACATCATCATCAGCTCCCA..... 9193
124 rpglyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
1914 .....TCGGTTCTCTAGTCTCTGCGCCCTAGTGCAA.....CTGCAA 9229
141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProG1 157
9230 TCCTTCTGCGCCCTGCTCTCTG..... 9253
157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
9254 .....CAGTGCTAGCTCAGGCCACCACTGCTCAGTGCATCTCACA 9295
seq_name: /SDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAS35759
seq_documentation_block:
ID AAS35759 standard; DNA; 31348 BP.
XX
AC AAS35759;
XX
DT
DE
DE Human cardiovascular system antigen genomic DNA SEQ ID No 1259.
XX
XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
OS Homo sapiens.
XX
XX WO200155321-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01340.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 09-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225271.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226686.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 21-SEP-2000; 2000US-0233085.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234997.
XX 25-SEP-2000; 2000US-0234998.
XX 26-SEP-2000; 2000US-0235484.
XX 27-SEP-2000; 2000US-0235834.
XX 27-SEP-2000; 2000US-0235836.
XX 29-SEP-2000; 2000US-0236327.
XX 29-SEP-2000; 2000US-0236367.
XX 29-SEP-2000; 2000US-0236368.
XX 29-SEP-2000; 2000US-0236369.
XX 29-SEP-2000; 2000US-0236370.
XX 02-OCT-2000; 2000US-0236802.
XX 02-OCT-2000; 2000US-0237037.
XX 02-OCT-2000; 2000US-0237038.
XX 02-OCT-2000; 2000US-0237039.
XX 13-OCT-2000; 2000US-0237040.
XX 13-OCT-2000; 2000US-0239935.
XX 13-OCT-2000; 2000US-0239937.
XX 20-OCT-2000; 2000US-0241808.
XX 20-OCT-2000; 2000US-0241809.
XX 20-OCT-2000; 2000US-0241826.
XX 01-NOV-2000; 2000US-0244617.
XX 08-NOV-2000; 2000US-0246474.
XX 08-NOV-2000; 2000US-0246475.
XX 08-NOV-2000; 2000US-0246476.
XX 08-NOV-2000; 2000US-0246477.
XX 08-NOV-2000; 2000US-0246478.
XX 08-NOV-2000; 2000US-0246523.
XX 08-NOV-2000; 2000US-0246524.
XX 08-NOV-2000; 2000US-0246525.
XX 08-NOV-2000; 2000US-0246526.
XX 08-NOV-2000; 2000US-0246527.
XX 08-NOV-2000; 2000US-0246528.
XX 08-NOV-2000; 2000US-0246532.
```


PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2000US-0254097.
 PR 05-JAN-2001; 2000US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.

XX PA
 XX PI
 XX PX
 XX DR
 XX WPI; 2001-451930/48.

XX New cardiovascular system related polynucleotides and polypeptides,
 XX useful for diagnosing, treating and/or preventing disorders of the
 XX cardiovascular system

XX Claim 1; SEQ ID No 1259; 674pp; English.

XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
 XX the cardiovascular system antigen polypeptides of the invention.
 XX Cardiovascular system antigens and their associated polynucleotides are
 XX useful in the diagnosis, treatment and prevention of various types of
 XX disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
 XX chickens or sheep. A pathological condition can be determined by
 XX detecting the presence or absence of a mutation in a cardiovascular
 XX system antigen polynucleotide. The treatable disorders include autoimmune
 XX diseases such as rheumatoid arthritis, hyperproliferative disorders such
 XX as neoplasms of the breast or liver, cardiovascular disorders such as
 XX cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
 XX nervous system disorders such as Alzheimer's disease, infections caused
 XX by bacteria, viruses and fungi, ocular disorders such as corneal
 XX infection, endocrine disorders such as premature labour and infertility,
 XX gastrointestinal disorders such as Crohn's disease, renal disorders such
 XX as glomerulonephritis and respiratory disorders such as asthma and
 XX pleurisy. The polypeptides can also be used to aid wound healing, to
 XX prevent skin aging due to sunburn, to maintain organs before
 XX transplantation, to regenerate tissues and in chemotaxis.
 XX Note: The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.

alignment_scores:
 Quality: 89.50 Length: 183
 Ratio: 1.041 Gaps: 10
 Percent Similarity: 46.995 Percent Identity: 25.683
 alignment_block:
 US-09-674-779-2 x AAS35759 ..
 Align seg 1/1 to: AAS35759 from: 1 to: 31348
 10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAla..... 22
 ||| :::::||||| :::::||||| :::::||||| :::::|||||
 8870 ACCTGTTTGTCTCACTGGGCTCTCTAGCATGCTCTTCCTGGCTGGTG 8919
 23
 8920 TGGGCTCTGCTTTTCTGCCCAAGCCCTGGGGTACCTGTT 8969
 27 snProGlnValSerProIleLysThrProSerValLeuIleThr..... 41
 :: ::| ||||| ||| :::::|||||
 8970 CCTTTTCACTCTCCAGGTGATCTCATCATGCTCCACACCATGAGC 9019
 42LysAspLysIleGlyAspHisThrHisGluHisAspGluSerVa 57
 :: ::| :::::||||| :::::|||||
 9020 CCACAGCCAGGCTCTCTGATCCATACCTGATCAC.....CT 9057
 57 lSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH 74
 ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
 9058 TTACACATGTCCTTCAGGCGGCCCTCAGACT..... 9091
 74 lAlaThrLysGlnGluValAlaArgTyrGlnAlaTyrLeuGlnSerArg 90
 ||||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
 9092CAAGCT...GTCCAGAGTAGA 9109
 91 LeuGlyAsnTyrLeuProMetSerGlnLeuLeuThrThrAlaArgse 107
 ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
 9110 GTTGGC...CATGGCCCACTTTACCAGATGTCACACCTGAACTCG 9156
 107 rTrpGlnAlaCysGlyHisGluProTyrGlnLeuProGluHisLeuT 124
 ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
 9157 AGTTCTTTCCTCCACCACATCCATCCAGCTCCCA..... 9193
 124 rPglyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArg 140
 ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
 9194TCGGTCTCTAGTCTCTGCCCTAGTGCAA.....CTGCAA 9229
 141 GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGl 157
 ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
 9230 TCCCTTTCCTCTGCTCTCTCTGTG..... 9253
 157 uLeuAsnGlnCys...AlaGlyGlyAlaAlaMetSerLysHisLeuThr 172
 ||||| ||| ::| ::| ::| ::| ::| ::| ::| ::|
 9254CAGTGCCTAGCTAGGCCACACTGCTCAGTCACTCACA 9295
 seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK79227
 seq_documentation_block:
 ID AAK79227 standard; DNA; 31348 BP.
 XX
 AC AAK79227;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34039.
 DE
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200157182-A2.
 XX
 PD 09-AUG-2001.

us-09-674-779-2.std.rng

Tue Sep 17 07:27:32 2002

```
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220563.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226968.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231124.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232168.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232599.
PR 14-SEP-2000; 2000US-0232600.
PR 14-SEP-2000; 2000US-0232601.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234273.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX
```

PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT	useful for preventing, diagnosing and/or treating cancers and
PT	metastasis -
XX	
PS	Disclosure; SEQ ID NO 34039; 3071pp + Sequence Listing; English.
XX	
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting
CC	the nucleic acids into a host cell and culturing the cell to express the
CC	protein. (I) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/hematopoietic-related diseases, especially
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC	to AAK87694 represent human immune/hematopoietic antigen genomic
CC	sequences from the present invention. AAK54942 to AAK34950 and AAM92169
CC	represent sequences used in the exemplification of the present invention.
XX	
SQ	Sequence 31348 BP; 7245 A; 7873 C; 8030 G; 8200 T; 0 other:

141	GlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProG1	157
9230	TCCTTCTTGCCTCGTCCGTGG.....	9253
157	uLeuAsnGlnCys....AlaGlyGlyAlaAlaMetSerLysHisLeuThr	172
9254	: : : : : : : :	
	9254CAGTGCTAGCTCAGGCCACCACTGCTCAGTCATCTCACA	9295
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.AAD10354		
seq_documentation_block:		
ID	AA010354 standard; cDNA; 2479 BP.	
XX	AA010354;	
XX	24-SEP-2001 (first entry)	
XX	Mouse tumour necrosis factor receptor/osteoprotegerin-like cDNA.	
XX	Mouse; cytostatic; tumour necrosis factor receptor/osteoprotegerin;	
KW	TNFr/OPG-like protein; antisense inhibitor; gene therapy; ss.	
KW	Mus sp.	
OS		
XX		
EH	Key	Location/Qualifiers
FT	CDS	91..1401
FT		/*tag= a
FT		/product= "Mouse TNFr/OPG protein"
FT	sig_peptide	91..183
FT		/*tag= b
FT	mat_peptide	184..1398
FT		/*tag= c
FT		/product= "Mouse mature TNFr/OPG protein"
XX	WO2001144472-A1.	
XX	21-JUN-2001.	
XX	14-DEC-2000; 2000WO-US33858.	
XX	16-DEC-1999; 99US-0172306.	
XX	(AMGE-) AMGEN INC.	
XX	Jing S, Welcher AA, Fox GM, Shu J, Boedigheimer MJ, Bennett BD;	
XX	WPI; 2001-451665/48.	
DR	P-PSDB; AAE05518.	
XX		
XX	New tumor necrosis factor receptor / osteoprotegerin-like	
PT	(TNFr/OPG-like) polypeptides useful for diagnosis and treatment of	
PT	associated disease -	
XX		
XX	Claim 1; Fig 2; 208pp; English.	
XX		
XX	The present sequence is a cDNA encoding mouse tumour necrosis factor	
CC	receptor/osteoprotegerin (TNFr/OPG)-like protein. TNFr/OPG-like prot	
CC	is useful for the treatment, prevention or amelioration of a medical	
CC	condition in a mammal resulting from decreased levels of TNFr/OPG	
CC	protein. Nucleic acid sequences of the present invention are used to	
CC	map locations of the TNFr/OPG genes and related genes. They are also	
CC	used as antisense inhibitors of TNFr/OPG expression. TNFr/OPG-like	
CC	protein and cDNA sequences are also used in gene therapy.	
XX		
XX	Sequence 2479 BP; 498 A; 711 C; 725 G; 545 T; 0 other;	

```

alignment_scores:
  Quality: 89.00
  Ratio: 1.035
  Percent Similarity: 45.745
  Length: 188
  Gaps: 10
  Percent Identity: 21.809
alignment_block:

```

```

US-09-674-779-2 x RAD10354/rev  ..
Align seg 1/1 to reverse of: RAD10354 from: 1 to: 2479

18 ValAlaCysSerAlaProIleProThrAsnProGlnVal..... 30
:|||||: :|||||: :|||||: :|||||: :|||||:
212 CTAGCTGGCTGCTCCTCTCCACAGAGCTCCAGAGCTTGAGGGCAGCAG 2163
31 .....SerProIleLysThrProSerVal. 38
|||||: :|||||: :|||||: :|||||: :|||||:
2162 GGTGGACTGAGGCTCTGCACAGAGAGCCCTGCCAAGTCAAGTCATTTT 2113
39 .....LeuIleThrLysAspLysIle.GlyAspHisHisThr..... 50
:|||||: :|||||: :|||||: :|||||: :|||||:
2112 TCAGTTTCATTAATAATAGCAGGTGAGGGCAGGCACATGCATTAACCC 2063
51 .....HisGluHisAspGluSerVal.....SerHisValGlyL 62
|||||: :|||||: :|||||: :|||||: :|||||:
2062 CTTCTGAAGGCATCACCATAATAGACATGCCATGCCATGGAGCCATGAAGGC. 2014
62 euGlnAlaHisPheGluThrTrpLeuGlnMethHisHisAlaThrLysGln 78
|||||: :|||||: :|||||: :|||||: :|||||:
2013 ..CAAGGCCCCAGAGCAGCAGCAAAACAGGCCAGCATAAAGCTACTATGGG 1966
79 GluValValArgTrpGlnAlaThrLeuGlnSerArgLeuGlyAsn..... 93
|||||: :|||||: :|||||: :|||||: :|||||:
1965 AGGCTAGCAAAAGTCAAGCTCCCCCACAACAGGACTCAGCCCCAGAAATCTGCA 1916
93 ..... 93

```

94TyrLeuProMetSerGlnLeu 101
 ||| ::: |||
185 GGATAAGCCCGAGTTGGAAACCAGGTACTCTGCCACCATAGAGGC TC 1816

102 LeuThr¹AlaArgSerTrpGlnAlaCys..... 111
 ::::: ||| ::::: ||| |||
 ::::: ||| ::::: ||| |||

```
cl-wiz      cl-progC]nIeuproProg]uHisiEit 124
```

112GlyHis...LeuProtyrGlnAspLeuPheGluAlaMetSer
1765 CCAGGGTCCAGACATGGGCACAGATGGCTAGTTCCCAAGCACACGCTT 1716
||| ||| : : : : :
124 rpGlyGlnIle...ValProThrLeuHisLeuTyrrGlnAspLeuLysSer 139
||| ||| ||| ||| ||| ||| ||| ||| : : : : :
1715 GGGGCCAGGTACAGTACCACTCCTCCACCAACACAGACCTCTAAGGGAC 1666

140 ArgGlyIleLeu 143
||||| : : : : :
1665 AGAGGGCTCCTA 1654

82 ArgTyrGlnAlaYrrLeuGlnSerArgLeuArgAsyLysLeuArgLys
::: :::: cccccc ||| cccccccccccacacggcgtcacgagatctccac
::: :::: cccccc ||| cccccccccccacacggcgtcacgagatctccac
::: :::: cccccc ||| cccccccccccacacggcgtcacgagatctccac

seq documentation_block:
ID: AB012149 standard; CDNA: 1362 BP.

26-MAR-2002 (first entry) 375

XX
XZ
WZ

Procellaria: developmental biology: cell signalling: insecticide: 416

Drosophila melanogaster.
XX
X
125 GlycyllevalprothrLeu..... 151

XX	27 SEP-2001
XX	
132HISLEUÝIG 133
132

23-MAP-2001, 2001WO-IIS09231].

22 JULY 2001 / 2002 AND 2003

```
135 InAspLeuLysSerArgGlyIleLeuProAlaAsn...ThrGlnIleArg 150
::||||| :::: :::: ::::
517 GAGATCTTCTCAGAGCAGCAGTACAGCTCCGACTTGTGCTGTGGT 566
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
```

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS30562

seq_documentation_block:

ID AAS30562 standard; DNA; 1580 BP.

XX AAS30562;

XX DT 21-NOV-2001 (first entry)

XX DE DNA encoding novel prostate gland antigen, Seq ID No 420.

XX Human; neurotrophic; cytotrophic; antiparkinsonian;
KW antianemic; dermatological; immunosuppressive; antinflammatory;
KW antiarthritic; antirheumatic; virucide; hepatotropic; nephrotrophic;
KW osteoplastic; prostate gland; prostatitis; adenocarcinoma; hair loss;
KW prostatic; malacoplakia; adenocarcinoma; benign prostatic hypertrophy;
KW hyperplasia; carcinoma; prostate neoplastic disorder; skin aging;
KW reproductive system disorder; autoimmune disorder; urinary system;
KW systemic lupus erythematosus; rheumatoid arthritis; cardiovascular;
KW blood-related disorder; hyperproliferative disorder; respiratory;
KW neurological disorder; endocrine disorder; inflammatory disorder;
KW liver disorder; wound healing; food preservative; ds.

OS Homo sapiens.

XX WO200155447-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01330.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.


```
||||:||||| :||||:||||| :|||||
769 GAGAGGAGGTAGAGGTGATTCAGGTGGGACCCAGAGATTCCTTT 818
215 rlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyA 232
819 TTCCTGCCCGCCCTGCACCTGAATGCCAG.....AAATGGGTT 859
232 laGlnPheSerGlnThrAsnSerIleCysArgHisValLeuProLys 247
860 TA....TTTCACGAG.....TCCTTCTGTCCAC...CTACACAGA 894

seq_name: /SIDS1/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA106616
seq_documentation_block:
ID AA106616 standard; DNA; 1580 BP.
XX AC AA106616;
XX DT 21-NOV-2001 (first entry)
XX DE Human reproductive system related antigen DNA SEQ ID NO: 9304.
XX KW Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy; ds.
XX OS Homo sapiens.
XX PN WO200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
```

17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249300.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen -
is used in preventing, treating or ameliorating a medical condition -
Disclosure; SEQ ID NO 9304; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders,
including cancer. The present sequence is a genomic sequence encoding a
protein of the invention.

Sequence 1580 BP; 353 A; 467 C; 401 G; 359 T; 0 other;

alignment_scores:
Quality: 88.50 Length: 267
Ratio: 0.770 Gaps: 15
Percent Similarity: 43.071 Percent Identity: 22.846

alignment_block:
US-09-674-779-2 x AAL06616 ..

Align seg 1/1 to: AAL06616 from: 1 to: 1580

20 CysSerAlaProIleProThrAsnProGlnVal.SerProIleLysThrP 36
236 TGCCCTTACCCACCAACACACAGGAGTTCAGATTATGGCGAA 285
36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHis... 51
286 TGGCCATTCTATACAAATCAAGGGTGACCAAGCACCACCTCCCGCA 335
52GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPh 66
336 ATGCTAAGGCACCCACAGAGGAGGAGGATAGTGTCTCCAAATTTT... 383
66 eGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgT 83
383 383
83 yrGlnAlaTyrLeuGlnSerArg...LeuGlyAsnTyrLeuProMet 98
384GCATTACACTCAAGGGTCTGGAGTCATATCTAGAAAAGGTC 425

99 SerGlnLeuLeuThrThrAlaArg.....SerTrpGlnAlaCysG 112
426 CAGATGGCTTCAATCCTAGCCAGACCCACACCTTCATCTCTGGG 475
112 yHisGluProTyrGlnLeuProGluHisLeuTrpGlyGlnIleValP 129
476 ACACAGCTCTCTAAACTCCCATATGACATCTCTGG..... 512
129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145
513 ..AGCCTCAGACTCCTTCAGGGA..... 533
146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
534CCTGTCTGAGGCTCCACACACCTCTGCTG.....GC 568
162 aglyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIle 179
569 AGGTGGGCTGTCAGGAGGAGGAGGAGCACCTGGGCCACCTGTGTCGGT 618
179 rpVal..... 180
619 GGTGTCTGAGTCCCTCTGTGTGCTGTAGAGAGCATCTGCCTTCACC 668
181ProAspLeuGluIleLe 186
669 ACCTCAGGCTGGAGAGTGAAGACACGGTAGCCCTGAGTATGTCTGTACC 718
186 sSerGlnAlaLeuTyrGlnLeuGlnAsnArgLeuCysGlnTyrTrpLeu 202
719 AAGCAGGGGTGCTCTACAGTAGATCCGAGGCCAGAGCCAGTCTATACA 768
203 GluHisGlyGluAsnGlnAsnPheGlyLeuGly.....LeuTy 215
769 GAGAGGAGGTAGTAGAGGTCTATCCAGGTGGAGCCACAGAGATTCTTT 818
215 rAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyA 232
819 TTCTTTGCCCGCCCTGCAATGGATGCCAG.....AAATGGGGTT 859
232 laGlnPheSerGluThrAsnSerIleCysArgHisValLeuProLys 247
860 TA...TTTCACGAG.....TCCTTCTGTCCAC...CTACCACGA 894

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.AAS80217

seq_documentation_block:
ID AAS80217 standard; cDNA; 900 BP.

XX AAS80217;
XX AC AAS80217;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #16021.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;


```

163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp..... 177
    ||||| ::::::::::: ::::::::::: |||
402 TGTGTACAGCGCTGACACCGCAATGGTGACACCGTTTCAGACAAACGTG 451
    ::::::::::: ::::::::::: |||||
178 .....IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
    ||||| ::::::::::: |||||
452 CAAACCCACCTGTGACGCCACGCCAACATCCGC.....GCCGGTAT 495
    ::::::::::: :::::::::::
192 GluLeuGlnAsn..... 195
    :: |||||
496 AAAGCACAGGGAATTGAGTTGCCTGCGGTAGTCGCGCTGGACTGGATAA 545
    :::::::::::
196 .....ArgLeuCysGlnTyr.....TrpL 202
    ||||| ::::::::::: |||
546 CCCAATGGGCCCATCATCGGATTCGCTGCGCGGCTATGGCGGGCTTATT 595
    :::::::::::
202 euGluHisGlyAluAsnGlnAsnPheGlyLeuGlyLeu...TyrAlaThr 217
    ||||| ||||| ::::::::::: |||||
596 TGCCTCATGGTACGACGCCCGCATTCGCGCATTGGCATGGCGGTAACTTCT 645
    :::::::::::
218 GlyAlaIleHisLeu 222
    ||||| ::::::::::: |||
646 GGCTGTATTGCTCTG 660

```

Quality:	88.00	Length:	255
Ratio:	0.807	Gaps:	11
Percent Similarity:	42.745	Percent Identity:	20.784

US-09-674-779-2 x AAH84500

8 PheIleThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProI 24

TG TGGCCTTACCGGC /7

78 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
79 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
80 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
81 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
82 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
83 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
84 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
85 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
86 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
87 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
88 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
89 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
90 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
91 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
92 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
93 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
94 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
95 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
96 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
97 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
98 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
99 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||
100 AAAAGCCAAACACCTGGGGCGGTGGCG
:::|||||

41 hrLysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57

---CGGAGGGCCTTTT...CATGTGGTGAAAAATACGGTGGT... 162

[illegible]

74 sAlaThrLysGlnGluValArgTyrGlnAlaTyrLeuGlnSerArgL 91

.....TCTGGGCTTCGTACAGGCTAACC 229

224 CCGCGTTGATCCTTACGTACCGCGCGCGCGGCGGCGGTCTTACC

106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProGluHi 122

1997

268ATCCCGTTGCAAAACCTACTTCCGATCCCCC 303

139 erArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyArgAsn 155

[illegible]

||| ||| :::
352 CCGGGTAAAAATTCCGTACCTCATAATCCCAAGGTCGCATTTA

100

XX
AC AAS90352:

XX

DE DNA encoding

KW Human; ch
KW food; ch

OS Homo sapi.

XX
PN WOZ001/50XX
DE 30-MAR-30

PR 31-MAR-20

PA (HYSE-) H:

PI Drmanac R
yy

DR P-PSDB; AN

PT New Isolation diagnostic

PT biodiversity:

XX

CC polypeptide
CC polymerase

CC polynucleotides for identification

 $(II) \cdot (II)$

CC a food sup

XX
SQ Sequence 1014 BP; 283 A; 234 C; 232 G; 265 T; 0 other;

alignment_scores:
Quality: 88.00 Length: 211
Ratio: 0.936 Gaps: 13
Percent Similarity: 44.550 Percent Identity: 27.962

alignment_block:
US-09-674-779-2 x ABL03309 ..

Align seg 1/1 to: ABL03309 from: 1 to: 1014

48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
||||| :|||:||||| :|||:
402 CATCATATACCCATCATGAGGAGATGCCAGCAC..... 437

64 aHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
||| :|||:||||| :|||:
438TCG...AGTTTCCATCATGAAATCGAACCTCACCGG 471

81 aArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
||||| :|||:||||| :|||:
472 TTCATATCGGTGATATTTAGCTCAGCAGAGTGGCTATTA. ATCCCG 520

98 MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisG1 114
||||| :|||:||||| :|||:
521 ATGTCT.....TCATGCCGGATCTGCATGCTCT 549

114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThrL 131
:|:|:| :|||:||||| :|||:
550 TTGCGTGGTGGAGCCGAGTACTTGTGG.....ATACATAT 597

131 euHisLeuTyrGln...AspLeuLysSerArgGlyIleLeuProAlaAsn 146
||||| :|||:||||| :|||:
588 TAACAACCTATCAGTAAGACTTAAAGAG..... 616

147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG1 163
||||| :|||:||||| :|||:
617CCAAACCCCAAGATAGTTCAA..... 637

163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpV 180
||||| :|||:||||| :|||:
638 .GGGTACGCTATG.....C 650

180 alProAspLeuGluIleLysSerGlnAlaLeu..... 190
:|:|:| :|||:||||| :|||:
651 TCCCAAGATTACCAAGTAAATACAGAAATATCTGTCGAGCTCAGTG 700

191TyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHi 204
||||| :|||:||||| :|||:
701 CCATTGAAATATACACTGCTATCGAGCTTGTGA.....TGGCTACAAA 745

204 sGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGly...AlaI 220
||||| :|||:||||| :|||:
746 ACCTTACAACCAACTTTTGGCAACGACAGTCAGTTGCTGGATGCCGA 795

220 leHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPheSerGlu 236
||| :|||:||||| :|||:
796 TT.....ACCGAGAAGAGAGTTCCTCGT 818

237 ThrAsnSerIleCysArgHisValLeuProLys 247
:|:|:| :|||:||||| :|||:
819 CCAAGAACCTGGTGC...CATCAGCTACCTAAG 848

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAS46271

seq_documentation_block:

XX ID AAS46271 standard; DNA; 12368 BP.
XX AC AAS46271;
XX

DT 18-DEC-2001 (first entry)
XX
DE DNA encoding novel mar regulated protein (NMR) #40.
XX
KW mar regulated polypeptide; NMR; microbial infection; antibacterial; ds.
XX
OS Escherichia coli.
XX
PN WO200170776-A2.
XX
PD 27-SEP-2001.
XX
PF 08-MAR-2001; 2001WO-US07478.
XX
PR 10-MAR-2000; 2000US-188362P.
XX
PA (TUFT) TUFTS COLLEGE.

Levy SB, Barbosa TM, Alekshun MN;

WPI; 2001-602769/68.
P-PSDB; AAU29372.

Identifying compounds that modulate a newly identified mar regulated polypeptide activity, useful as antimicrobial compounds, involves contacting the polypeptide with a test compound -

Disclosure; Page 458-466; 526pp; English.

The invention relates to a method of identifying compounds that modulate a newly identified mar regulated (NMR) polypeptide activity. The method comprises contacting an NMR polypeptide with a test compound under interaction conditions, determining the ability of the compound to modulate the activity or expression of the polypeptide, and selecting the modulators. NMR nucleic acids and polypeptides are used in the treatment of microbial infections, and in screening for modulators of NMR expression and activity. These modulators can be used to reduce the infectivity of a microbe on a surface, and the virulence of a microbe in a subject suffering from an infection. AAS46232-AAS46278 represent Escherichia coli NMR coding sequences of the invention.

SQ Sequence 12368 BP; 2959 A; 3148 C; 3276 G; 2985 T; 0 other;

alignment_scores:
Quality: 88.00 Length: 255.
Ratio: 0.807 Gaps: 11
Percent Similarity: 42.745 Percent Identity: 20.784

alignment_block:

US-09-674-779-2 x AAS46271/rev ..

Align seg 1/1 to reverse of: AAS46271 from: 1 to: 12368

8 PheLeuThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProI1 24
:|:|:| :|||:||||| :|||:
8659 TGGTACAGCTTTTATACGTCGCGCTGCCGTGGCGCTACCGC 8610

24 eProThrAsnProGlnValSerProIleLysThrProSerValLeuI1 41
:|:|:| :|||:||||| :|||:
8609 AAAGCCCAACACCTGGCGCTGCCG...CCAGCGGCGAGTCGTCTGTTG 8563

41 hrLysAspLysIleGlyAspHisThrHisGluHisAspGluSerVal 57
:|:|:| :|||:||||| :|||:
8562 CGCAAAACAAATTT.....CATGTGCTGGAATGACGGTGGT... 8525

58 SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisH1 74
:|:|:| :|||:||||| :|||:
8524TCTCTGGAAGC 8514

74 sAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgL 91
:|:|:| :|||:||||| :|||:
8513 CATGCCCAAAATATACACGTCGCTTCTCGCTCTGTTACAGGCTAAC 8464

```

91 euGly.....AsnTyrLeuProMetSerGlnLeuLeuThrAla 105
   |||  |||:|||||  :||||:|||||
8463 CCGCGTGTGATCCTTACGTACCGCGCGGAGCGGTAAACG..... 8420
106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProGluHi 122
8420 ..... 8420
122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLys 139
   :|||  |||  |||
8419 .....ATCCGTTGCAACCCCTACTTCCAGATGCGCGC 8386
139 eArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 155
   :|||:|||||  :||||:|||||
8385 GCGAAGGCAPTGTGATCAACATTCGCGAGCTGGCTCTCTATTACTACCG 8336
156 ProGluLeuAsnGlnCysAla.....G1 163
   |||  |||:|||||  :|||
8335 CCGGTTAAATTCGGTAACCGTTATCCCAATAGGTATTGGTCAGTTAGG 8286
163 yGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp..... 177
   |||  :||||:|||||  :||||:|||||
8285 TGGTGACACGCTGACACCGCAATGTGTACCAACCGTTTCAGACAAACG 8236
178 .....IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr 191
   |||:|||||  :||||:|||||  :||||:|||||
8235 CAAACCAACCTGGACGCCCAACGCGCAACATCCGC.....GCCCGTTAT 8192
192 GluLeuGlnAsn..... 195
   :|||  |||:|||||
8191 AAACACACAGGAATTGAGTTGCTGCGGTAGTGCAGGCTGACTGGATAA 8142
196 .....ArgLeuCysGlnTyr.....TrpL 202
   |||:|||||  :|||
8141 CCCAATGGCCATCATCGCATTCGTCTGCGGCTATGCGCGGTATT 8092
202 euGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu...TyrAlaThr 217
   |||  |||  |||  :||||:|||||  :||||:|||||
8091 TGCTTCATGTGTACCAACCGCATTCGCGATTGCGATTCGCGGTAAAGTTCT 8042
218 GlyAlaIleHisLeu 222
   |||:|||||
8041 GGCTGTATTGCTCG 8027

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA82625
seq_documentation_block:
ID   ABA82625 standard; DNA; 66933 BP.
XX
AC   ABA82625;
XX
DT   25-JAN-2002 (first entry)
XX
DE   Human HBM gene region b200e21-h_contig4.
XX
KW   Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW   sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW   antisenese therapy; vaccine; bone disorder; Paget's disease;
KW   sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177327-A1.
XX
PD   18-OCT-2001.
XX
PF   21-JUN-2000; 2000WO-US16951.
XX
PR   05-APR-2000; 2000US-0543771.
XX
PR   05-APR-2000; 2000US-0544398.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
```

```

XX
PI   Carulli JP, Little RD, Recker RR, Johnson ML;
DR   WPI; 2001-657171/75.
XX
PT   New high bone mass (HBM) and Zmax1 genes and proteins useful for
XX   modulating bone mass for the treatment of e.g. osteoporosis -
XX   Claim 51; Page 308-350; 443pp; English.
XX
CC   The present invention describes the human Zmax1 gene and the high bone
CC   mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
CC   HBM genes have osteopathic activities. The genes can be used in gene
CC   therapy, antisenese therapy and in the production of vaccines. They
CC   can be used in the diagnosis and treatment of bone disorders including
CC   osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC   dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;

alignment_scores:
Quality: 88.00      Length: 248
Ratio: 0.880       Gaps: 15
Percent Similarity: 40.323 Percent Identity: 26.210

alignment_block:
US-09-674-779-2 x ABA82625/rev ..
Align seg 1/1 to reverse of: ABA82625 from: 1 to: 66933
23 ProIleProThrAsnProGlnVal.....SerProIleLy 34
   |||  |||:|||||  :||||:|||||
39344 CCCACACACACACACCCCGAGATGGAAGGCTGCTCCTCCCTCCCATACA 39295
34 sThrProSerValLeuIleThrLysAspLysIleGlyAspHisThrH 51
   |||  |||  |||  |||:|||||
39294 CACACACACACCA.....CACACACACACACAC 39266
51 sGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
   ||  |||  :||||:|||||  :|||
39265 ACACACACACACACACACACACACACACACACAGCCTTGCT..... 39225
68 ThrTrpLeuGlnMetHisHis.....AlaThrLysGlnGluVa 80
   |||  |||  |||  |||  :||||:|||||
39224 GGGCAGCTGCAAGACATCATCTTCATCTGGACACGAAGGATCATGT 39175
80 lValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProp 97
   ||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
39174 G..... 39174
97 roMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGly... 112
   |||  :||||:|||||  |||  |||:|||||
39173 .....TCTTTGTAGATAACAACATCCACATCCACTACAGAGCTGTGACCC 39129
113 .....HisGluProTyrGlnLeuProGluHi 122
   :|||  :|||  |||  |||  |||  |||
39128 TAGGGGGCATGTGTCGAAGCCAATGTGAACAATTCCTCTCTAAAGCCA 39079
122 s...LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuL 138
   |||  |||  |||  |||  |||  |||
39078 CTAGCTGTGG.....ACCTGGACTGGTACAGC..... 39051
138 ysSerArgGlyIleLeuProAlaAsnThrGlnIleArg.SerValTyrAr 154
   |||:|||||  |||  |||  |||  |||
39050 .....TCCCGGCTGAAGACAAGCAAGATTCAGCTAC... 39017
154 gAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisL 171
   |||  |||:|||||  :|||
39016 ....CCTGGCCAAATCAGGCTCAGCGTGGAGG..... 38987
171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
```


241 sArgHisValLeuPro 246
|||
396 CCTTCACTGTCCCCA 381

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK80624

seq_documentation_block:

ID AAK80624 standard; DNA; 5858 BP.

XX AC AAK80624;

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35436.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

PN WO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184564.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225271.

PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241185.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

[illegible]

```

alignment_scores:
  Quality: 87.00      Length: 162
  Ratio: 1.338        Gaps: 8
  Percent Similarity: 40.123  Percent Identity: 23.457

alignment_block:
US-09-674-779-2 x AAK80625 ..

Align seg 1/1 to: AAK80625 from: 1 to: 5866

20 CysSerAlaProIlePro.....ThrAsnProGlnValSerProI1 33
||||: |||:||||| |||: |||: |||
2329 TGTGAGTAGCAGTCCATTCATTTTGACAAACACACAGGCCACCCAGC 2378
33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHisH1 50
|||: |||: |||: |||: |||: |||: |||
2379 TCTGTCGCCAGGAGTG.....CACACACACACACACACACAC 2398
50 hrHisGluHisAspGluSerValSerHisValGly.....61
||||: |||: |||: |||: |||: |||: |||
2399 CACATGAGCATAGTAAATCCACAAACACACCGCGGTGGTAATGTT 2448
61 .....61
2449 ATGCTCATTTTACAGAGGAGGAAATTTGAGTTCAGAGAGGCAAGAC 2498
62 ....LeuGlnAlaHisPheGluThrTriLeuGlnMetHisAlaThrL 77
|||: |||: |||: |||: |||: |||: |||
2499 TTACCTGGGTCCCATATCCCATGCTGGCAAGTGCACACACCAACCTG 2548
77 yScGlnGluValVal.....ArgTyrGlnAlaTyrLeuGlnSerArg 90
||||: |||: |||: |||: |||: |||: |||
2549 TCCAAAACTATCCAGCCAGGGAAGCTGTCTTACCTGGAGGAGA 2598
91 LeuGlyAsnTyrLeuProMetSerGlnLeu.....ThrAla 105
|||: |||: |||: |||: |||: |||: |||
2599 GGTGGTGGTAGTCTGGGAGCAGGCGAGCGAGCTCATGGGCGAGTGC 2648
105 aArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu..... 118
||||: |||: |||: |||: |||: |||: |||
2649 AAGAGCCTGGTCTCGGGAACACACAGACCT...CAGCTCAATCCAGC 2695
119 .....ProProGlu.....121
2696 TCATCATCTGTGTGACTTTAGAAAATGACCCCTCTCTGGGACTCAGT 2745
122 ...HisLeuTrpGlyGlnIleValProThrLeuHis 132
|||: |||: |||: |||: |||: |||: |||
2746 TTTCACATGGAAGATGAGGATACCAATTCACAT 2781

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABL25734

seq_documentation_block:
ID ABL25734 standard; DNA: 3864 BP.
XX
AC ABL25734;
XX
DT 26-MAR-2002 (first entry)
XX
Drosophila melanogaster genomic polynucleotide SEQ ID NO 28675.
XX
Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
Drosophila melanogaster.
OS
XX
W0200171042-A2.
XX
27-SEP-2001.
XX
23-MAR-2001; 2001WO-US09231.
XX

```

```

PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 28675; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3864 BP; 1078 A; 764 C; 842 G; 1180 T; 0 other;

```

```

alignment_scores:
  Quality: 86.50      Length: 146
  Ratio: 1.153        Gaps: 8
  Percent Similarity: 51.370  Percent Identity: 26.027

alignment_block:
US-09-674-779-2 x ABL25734/rev ..

Align seg 1/1 to reverse of: ABL25734 from: 1 to: 3864

7 TyrPheIleThrLeuIle.....SerSerMetLeuValAl 19
||||: |||: |||: |||: |||: |||: |||
1572 TATTACTCCTCAAAATCTAGTTTACTTACCCTGGAAGTCAACCCATCATAGC 1523
19 aCysSerAlaProIleProThrAsnPro.....GlnValSerProIle 34
|||: |||: |||: |||: |||: |||: |||
1522 CATCGAATATTCGCGCGGTGAACCTTTTAAAGTAGTCCATTACCTCA 1473
34 yThrProSerVal.....LeuIleThrLysAsp.....43
||||: |||: |||: |||: |||: |||: |||
1472 AAACCTCTCTATAGGATAGTACCCCTCGTTTCGAAAGATTACTTCGTA 1423
44 .....LysIleGlyAspHisH1Thr...H1 51
||||: |||: |||: |||: |||: |||: |||
1422 GTCTGAGGATTTCATCGTTAGCACCAAAATCGTTCTCTGCACACTGGCA 1373
51 sGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluT 68
|||: |||: |||: |||: |||: |||: |||
1372 TCCGAAAGATATTCATATTTGCGTTTCATATCGAGAAACATTTCTGG 1323
68 hrTrpLeuGlnMetHisAlaThrLysGlnGluValValArgTyrGln 84
|||: |||: |||: |||: |||: |||: |||
1322 ACCGCAAAATTCCTATAGACCTTAAAGAGAGGATTTCAACAAAGATCA 1273
85 AlaTyrLeuGln.....SerArgLeuGlyAsnTyrLeuPro.P 97
|||: |||: |||: |||: |||: |||: |||
1272 ACAATTCTCGATGAGAGAATACTGGAAAGCTTTCTCTTCACTTACCGCC 1223
97 roMetSerGlnLeuThrThrAlaArgSerTrp...GlnAlaCysGly 112
|||: |||: |||: |||: |||: |||: |||
1222 CGGCAGAAAAGCTCTGATCAATTTGCAATTTACTGGTCCCGGAATATCAG 1173
113 HisGluProTyrGlnLeuProProGluHisLeuTrp 124

```


PA (HARD) HARVARD COLLEGE.

XX Shaprio R, Vallee BL;
XX WPI; 1988-331302/47.
DR P-PSDB; AAP80646.
XX Inhibitors of angiogenin, useful for inhibiting tumour growth -
PT include human placental RNase inhibitor and active polypeptide
PT segments
XX Disclosure; ; p; English.
XX cDNA sequence isolated from human placental library using probes
CC AAN81071 to AAN81082. Tryptic digestion of the PRI protein encoded by
CC this sequence produces peptide fragments possessing angiogenin
CC inhibitory activity.
CC Disorders associated with neovascularisation such as rheumatoid
CC arthritis and Kaposi's sarcoma are treated by admin of these
CC inhibitory peptides.
XX
SQ Sequence 1698 BP; 316 A; 535 C; 535 G; 312 T; 0 other;

alignment_scores:

Quality: 86.00 Length: 268
Ratio: 0.694 Gaps: 15
Percent Similarity: 46.269 Percent Identity: 25.373

alignment_block:

US-09-674-779-2 x AAN81083/rev ..

Align seg 1/1 to reverse of: AAN81083 from: 1 to: 1698

17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProII 33
1222 CTGGCCAGGCCCTGGCAGAGCTCCCGCAGCCGCGCAGCTCCAGCCTGT 1173
33 eLysThrProSerValLeuIleThrLysAspLysIleGlyAspHis1st 50
1172 TGTGCTTATCTAGCTCCAGGACAAACCTGTCTGGGCCAGCAC.... 1127
50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
1126TGAGCTGAAGTGGGAGCAGCAGCGCGCTGT 1097
67 GluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTy 83
1096 GAAGCT...GCAGGACTTCACCCAGCAGGACTCCAGCTGGCAGCCAGGT 1050
83 rGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMet.... 98
1049 CCAGCAGGCTCTCACAGCAGTGGGACCCCTCATCCCCAGCTCGTTG 1000
99SerGlnLeuLeu.....ThrAla 105
999 CGCGCCAGGCTGAGCTCTCTCAGGCTCTCTCTGGCCCTGAGGACACGGCA 950
106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro.... 120
949 CAGATCCCGCAGCCCTGGCAGTGTGCCACACTCCAGATCCACAGGG 900
121GluHisLeuTrpGlyGlnIleValProT 130
899 TCCTGAGCTGGAGCTGGGTGGAGCAGCCCTGGGACAGCTC...GCC 853
130 hrLeuHisLeuTyrClnAspLeuLysSerArgGlyIleLeuProAlaAsn 146
852 ATGCCACATCACCAGCTGTGTGCTGCCAGGCGCAGCTCCCCACGGA 803
147ThrGlnIleArgSerValTyr..... 153
802 GGCCTTGGAGGCCAACATGCCGACAGCTCCCGCAGTGTCTGATGCA 753

154ArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla.... 165
752 CACCGCAGCTCTCCAGCTTGAGCGCTCCAGCTGCAGGGGAGTCTCTTC 703
166 ..AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
702 AGGCCCTGGCAGCACACGACGACGACGACCTCATTTGTTGCT 653
182 AspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCy 198
652 AACCGTGAGCTC.....CTTGAAGTCCGCGCTGG 624
198 sGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
623 CCT.....GAGCAGCGAGGCCAG.....GGGCTCGCAGCTG 592
215 yAlaThrGlyAlaIleHisLeuAspThrGlnGlyPhe.ArgLys...Tr 230
591 GCAGCCGAGAGGCTGCAATACTCCAGCTGCAGCTTTTCCAGGCGGCACTG 542
230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisValLeuProL 247
541 GGGGTCCAGG...AGTCTCTCGAGAGCAGCTGCAGGCCCGCATCCCCCA 495
247 ys 247
494 AG 493

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.AAS75095

seq_documentation_block:

ID AAS75095 standard; cDNA; 1998 BP.

AC AAS75095;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #10899.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG10908.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 10899; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

```

31 erProLeuLys.....ThrPro 36
   ||| |||
5957 TTCGAGCAAAACGAACTAAACACGCGCGGACTCTACTCGCGC 5908
   ||| |||
37 SerValLeuLeuThrLysAspLysIleGlyAspHis..... 48
   ||| |||
5907 TCCGCGCACACACTCGCGCGCGCGTCCGCTCACTCACTCGCGCACAA 5858
   ||| |||
49 .HisThrGlnAlaHisAspLysValSerHisValGlyLeuGlnAlaHis 65
   ||| |||
5857 ACATCGCACTCTCACAGCACACACTCTCACACACACACACACAC 5808
   ||| |||
65 ispheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValVal 81
   ||| |||
5807 ACACACACACATACACAC.....ACAAACCCGAAACCGCC 5770
   ||| |||
82 ArgTyrGlnAlaThrLeuGlnSerArgLeuGlyAsnTyrLeu..... 95
   ||| |||
5769 GCGACCTAACCCAAATACGACGCGGAAACAACTACATCAAAACAA 5720
   ||| |||
96 .....ProMetSerGlnLeuLeuThr..... 104
   ||| |||
5719 AAAAAACGCGCGCGCGCGCGGCAATCGACTCCACGACGCGGAC 5670
   ||| |||
105 .....AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
   ||| |||
5669 CACGAAACACCCCGCAACTCTTACGCCCCCAAAATCCCGGACGC 5620
   ||| |||
118 LeuProProGluHisLeuTyrGlyGlnIleValProThrLeuHisLeu 134
   ||| |||
5619 CCTCTAAATAAATACGAAACACGACGACCCCGACTCTAAACGCGCG 5570
   ||| |||
134 r...GlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIle 150
   ||| |||
5569 ATAAATAAACGATGACGACCGCGTAACCTCTCTCGAAATACTAATA 5520
   ||| |||
150 rgSerValTyrArgAsnPro 156
   ||| |||
5519 CG.....AAAAATCCA 5509
   ||| |||

```

seq_name: /SDSL/gcdata/hold-geneseq/geneseq-emb1/WA1996.DAT:AA11658

seq_documentation_block:

ID AA11658 standard; DNA; 22481 BP.

AC AA11658;

XX 16-JAN-1997 (first entry)

DE PEDF full length sequence and flanking sequences.

XX Pigment epithelium-derived factor; PEDF; neuronal cells; neurons;
 KW glial cells; gliastatic; gliosis; central nervous system; CNS;
 KW neurodegenerative disease; injury; neurotrophic; brain cells;
 KW Parkinson's disease; photoreceptor cells; retina; inhibition;
 KW proliferation; immunoassay; antibody; ageing; degenerative disease;
 KW ds.

OS Homo sapiens.

XX WO9533480-A1.

XX 14-DEC-1995.

XX 06-JUN-1995; 95WO-US07201.

XX 30-DEC-1994; 94US-0367841.

XX 07-JUN-1994; 94US-0257963.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Becerra SP, Chader GJ, Schwartz JP, Taniwaki T;

```

XX WPI; 1996-039966/04.
DR P-PSDB; AAR90287.
XX
PT Use of pigment epithelium derived factor - for enhancing neuronal
PT cell survival and inhibiting glial cell proliferation, useful, e.g.
PT in CNS cell culture or to treat neuro-degenerative diseases
XX
PS Disclosure; Page 100-122; 151pp; English.
XX
CC Pigment epithelium-derived factor (PEDF) has both neurotrophic and
CC gliastatic activity, making it useful in cases where neurons die
CC quickly and glia tend to proliferate (gliosis), e.g. in CNS cell
CC culture, in neurodegenerative diseases and in CNS injury. The
CC neurotrophic effect of PEDF is especially useful for enhancing
CC survival of neuronal cell cultures intended for use in
CC transplantation. These include cultures of human foetal brain cells
CC and neural retina and photoreceptor cells. The gliastatic activity
CC of PEDF can be applied to inhibiting glial cell proliferation in
CC certain tumours. Antibodies directed against PEDF can be used for
CC inhibiting PEDF activity or in an immunoassay for determining
CC levels of PEDF in fluid, cellular or tissue samples e.g for
CC determining ageing and/or other degenerative diseases.
XX
SQ Sequence 22481 BP; 5280 A; 5708 C; 6136 G; 5347 T; 10 other;

```

```

alignment_scores:
  Quality: 86.00      Length: 298
  Ratio: 0.662      Gaps: 14
  Percent Similarity: 43.624      Percent Identity: 21.141
alignment_block:
US-09-674-779-2 x AA11658/rev ..
Align seg 1/1 to reverse of: AA11658 from: 1 to: 22481

```

```

23 ProLeProThrAsnProGlnValSerProIleLysThrProSerVal... 38
   ||| |||
21479 CCTGCCCCACCCCACTCA.....CCACCAAGAGCGAAAGGGTCTT 21439
   ||| |||
39 .....LeuIleThrL 42
   ||| |||
21438 CAGACATACCTATCTCTCGAGGACTTGGTGACTTCGCTCGTAATC 21389
   ||| |||
42 ysAspLysIleGlyAspHisHisThrHisGluHisAspLysSerValSer 58
   ||| |||
21388 AGCTTCAGCTTGGGACAGTGGAGACCGCGCTGCACGGTCTTCAGTTC 21339
   ||| |||
59 .....HisValGlyLeuGlnAlaHisPheGluThrTriple 70
   ||| |||
21338 GTCATGTGATGAATGAATCGGAGGTGAGGCTCTCTCTATCAAGGTCA 21289
   ||| |||
70 uGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
   ||| |||
21288 AATTCGTGGTCACTTTCAGGGGCGAGGAGAGATGATCATCTGCTCG 21239
   ||| |||
87 euGlnSerArgLeuGlyAsnTyr..... 94
   ||| |||
21238 GTCAGGGGCGAGCTGGGCAATCTGACACAAAGTGAGAAAGCATGTGTTAGT 21189
   ||| |||
95 ...LeuProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAl 110
   ||| |||
21188 CCTTCAGACCCCAAGCGGGGAGACACACACAGCAGGAGC..... 21147
   ||| |||
110 aCysGlyHisGluProTyrGlnLeuProProGluHisLeuThrGlyGln. 126
   ||| |||
21146 .TGCCTTCCTCTCCCGTGCACGGGCC.....TGGTCTCGTTC 21110
   ||| |||
127 IleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIle 143
   ||| |||
21109 CTTTCATCCAGCCCTCTCTATTATTACAAATTCCTCCAGGTATAGCT 21060
   ||| |||

```


xx
pt
pt
pt
pt
pt
pt
xx

```

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI99682
seq_documentation_block:
ID   AAI99682 standard; DNA; 4411529 BP.
AC   AAI99682;
XX
XX
DT   15-JAN-2002 (first entry)
XX
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX
OS   Mycobacterium tuberculosis.
XX
XX
US6294328-B1.
XX
XX
PD   25-SEP-2001.
XX
XX
PF   24-JUN-1998; 98US-0103840.
XX
XX
PR   24-JUN-1998; 98US-0103840.
XX
XX
PA   (GENO-) INST GENOMIC RES.
XX
XX
PI   Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX
DR   WPI; 2001-647261/74.
XX
XX
CC   Evaluating strain variation of Mycobacterium tuberculosis, comprises
CC   determining the nucleotide sequence of the strain at positions in the
CC   genome corresponding to positions where M. tuberculosis strains CDC
CC   1551 and H37Rv differ
XX
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99682) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
XX
SQ   Sequence 4411529 BP; 738565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

```

alignment_scores:

```

Quality: 86.00 Length: 183
Ratio: 1.075 Gaps: 11
Percent Similarity: 43.716 Percent Identity: 24.044

```

alignment_block:

```

US-09-674-779-2 x AAI99682/rev ..
Align seg 1/1 to reverse of: AAI99682 from: 1 to: 4411529

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProth 26
||||| ||| :|||:|||||:||||| :||
4304162ACCTTGACCTCCCGATCGTGGTCTCCGCATAGCAGCGAGTCACCAT 4304113

26 rAsnProGlnValSerProIleIleYsthrProSerValIleuIleThrLysA 43
||||| :|||:|||||:||||| :||
4304112CACAGAGCTGATCGATCCCGGAGTGTTCGCGTCG ..... 4304079

43 sPLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59

```

```

:::||||| |||:||||| :|||:||||| :|||:||||| :|||:|||||
4304078...GAAATCCGCATCGCTATTTCGCGC.....ATCGACAC 4304046
60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAla 76
||| :|||:|||||:||||| :|||:||||| :|||:|||||
4304045CGCGGC.....AGCTATCTGCAGATGCATTTTCGCT 4304014
76 rIysGlnGlu.....ValValArgTyrGlnAlaTyr..... 86
||| :|||:|||||:||||| :|||:||||| :|||:|||||
4304013CGCGCAGCACCGCGCTTCGCGCGCCCTACACGAGCTCAACGATCCGA 4303964
87 ..LeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGlnLeuLeu 102
:::||||| :|||:||||| :|||:||||| :|||:|||||
4303963GCATGCAGCAGCATCGATGGCATTTTCTGTACACCGGAGCAG..... 4303923
103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
::: |||:|||||:||||| :|||:||||| :|||:|||||
4303922...GTTTCAGCAGCAGTGGGAGGATTGC..... 4303899
119 OProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
4303899.....
136 sPLeuLysSerArgGlyIleLeuProAlaAsn.....ThrGln 148
|||||:|||||:||||| :|||:||||| :|||:|||||
4303898.....CGCGTGGGATCGTCCCGCGCGATCCGACCGTGTGTGCAG 4303857
149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
||| :|||:|||||:||||| :|||:||||| :|||:|||||
4303856ATCCCGTCGTCAT...GATCCAGCGTGGCGCGGGTAAAGCAGGC 4303810
165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleThrValPro 181
||||| :|||:|||||:||||| :|||:||||| :|||:|||||
4303809CCG.....TCGCGTTCGCAATGTGGTCCCG 4303782

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ53602
seq_documentation_block:
ID   AAZ53602 standard; DNA; 915 BP.
XX
XX
AC   AAZ53602;
XX
XX
DT   21-MAR-2000 (first entry)
XX
XX
DE   Neisseria meningitidis ORF 280 partial DNA sequence SEQ ID NO:1153.
XX
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
XX
OS   Neisseria meningitidis.
XX
XX
PN   WO9957280-A2.
XX
XX
PD   11-NOV-1999.
XX
XX
PF   30-APR-1999; 99WO-US09346.
XX
XX
PR   01-MAY-1998; 98US-0083758.
PR   31-JUL-1998; 98US-0094869.
PR   02-SEP-1998; 98US-0098994.
PR   09-OCT-1998; 98US-0099062.
PR   09-OCT-1998; 98US-0103749.
PR   09-OCT-1998; 98US-0103794.
PR   09-OCT-1998; 98US-0103796.
PR   25-FEB-1999; 99US-0121528.
XX
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
XX
PI   Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI   Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI   Tettelin H, Venter JC;

```

XX WPI; 2000-062150/05.
 DR P-PSDB; AAY74840.
 XX
 XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 XX Claim 7; Page 636; 1453pp; English.
 XX
 XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254617 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria meningitidis* (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria meningitidis*, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 XX SQ Sequence 915 BP; 277 A; 292 C; 195 G; 151 T; 0 other;

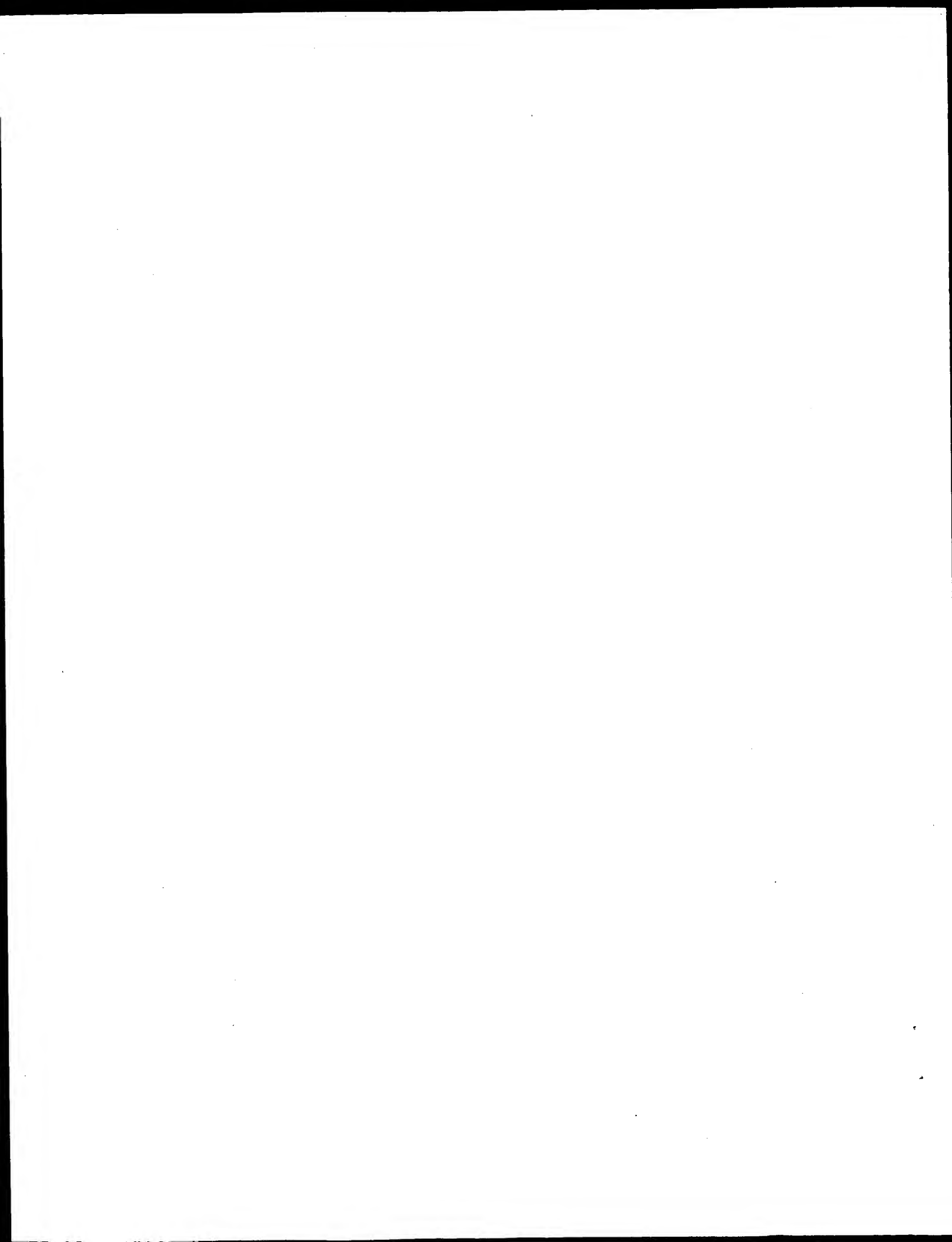
alignment_scores:

Quality: 85.50 Length: 215
 Ratio: 0.864 Gaps: 12
 Percent Similarity: 46.047 Percent Identity: 22.791

alignment_block:

US-09-674-779-2 x AA253602 ..
 Align seg 1/1 to: AA253602 from: 1 to: 915
 17 LeuValAlaCysSerAla.....ProileProth 26
 ||| |||||
 248 CTCCGATGTGCAACGCGCGTCAACAAAGCAAGTATCTCTATACCGAA 297
 26 rAsnProGlnValSer..ProlleLysThrProSerValLeuIleThrly 42
 :|||:|||||
 298 GCGACCAAGGCGATCCAAACCCCTCAA.....GCCGA 329
 42 sAspLysIleGlyAspHisThrHisGluHisAspGluSerValSerH 59
 :|||:|||||
 330 AGAAGAGGCGGACACATCACACACACATGATGACACGAGGACACC 379
 59 isValGlyLeuGlnAlaHisPheGlu.....ThrTrpLeuGln 71
 || :||| ||| :|||
 380 AC.....CATGACCGGCGAATATGACCCGACGCTGTGGAACGAC 420
 72MethHisAlaThrLysGlnGluValValArgTyr..... 83
 :|||:|||||
 421 CCGCTCTTATGTCGCTATGCCCAAAACGTTGCCAAAGCCCTGATAAA 470
 84GlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrL 95
 :|||:|||||
 471 GGCGCATCCGAGGCAAGATTATATCAACACGCTGGGCACTACC 520
 95 euProPrometSerGlnLeu..... 101
 :|||:|||||
 521 AAATGCACTCAAAAACCTGCACAGCGACGACCAAGCCGATTAATGCG 570
 102 LeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln.. 117
 :|||:|||||
 571 GTCCCTCTGCGCAACGCAAGTCTCTGACCGGCGACGATGCTTTTCCCTA 620
 118LeuProProGluHisLeuLeuTrpG 125
 :|||:|||||
 621 TATGGCAACAGTTACCATATCGAATTATCATCGCCCGCAAGGCGTGACGA 670
 125 lyGlnIleValProThrLeuHis.....LeuTyrGlnAspLeu 137

671 GCGAAGCGGAGCCTTCGCGCAAAAGTCGCGCCATCATCGACAAATC 720
 |||:|||||
 138 LysSerArgGlyIleLeuProAlaAsnThrGln.....IleAr 150
 ||| :|||:|||||
 721 AAACGCGAAGGCGATCAAGCGCTTTACCGAAACATCAAGGACACCGG 770
 150 gSerValTyrArg...AsnProGluLeuAsnGlnCysAlaGlyGlyAla 166
 ||| ||| ||| :|||
 771 TATGGTTGACCGTATCGCCAAAGAACCGGTGTCAACGTCAGCGGCAAC 820
 166 laMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpVal 180
 ||| ||| |||:|||||
 821 TGTATTCGCGACGCTCGGCAACGCGCCGACGACACCTACATC 864
 seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT:AAQ46887
 seq_documentation_block:
 ID AAQ46887 standard; DNA; 2188 BP.
 XX
 AC AAQ46887;
 XX
 DT 08-FEB-1994 (first entry)
 XX
 DE Sequence of chicken brain cDNA encoding acetylcholine receptor-inducing
 DE activity (ARIA) protein E18.
 XX
 KW acetylcholine receptor-inducing activity protein; ARIA; E18; ss.
 XX
 OS Gallus.
 XX
 FH Key
 CDS 172..972
 FT /*tag= a
 XX
 XX USS237056-A.
 XX
 PD 17-AUG-1993.
 XX
 PF 29-MAY-1991; 91US-0706872.
 XX
 PR 29-MAY-1991; 91US-0706872.
 XX
 XX (HARD) HARVARD COLLEGE.
 PA
 PI Fischbach GD;
 XX
 XX WPI; 1993-272184/34.
 DR P-PSDB; AAR40171.
 XX
 XX DNA encoding protein which co-purifies with acetylcholine
 XX receptor-inducing activity - used to obtain prods. for treating
 XX deficiencies in neural transmitter receptors, e.g. Alzheimer's
 XX disease
 XX
 XX Claim 2; Figure 2; 17pp; English.
 XX
 XX ARIA promotes the synthesis and accumulation of acetylcholine
 XX receptors (ACHRs) in muscle cells. The protein was isolated on the
 XX basis of its ability to increase the rate of insertion of ACHRs into
 XX the surface membrane of chick myotubes. Oligos with sequences
 XX corresp. to the sequence of the protein were used to amplify from
 XX chicken brain cDNA a 34-nucleotide sequence encoding 11 AAs from
 XX ARIA. The 34-nucleotide sequence corresp. to the 3' nucleotides of
 XX the 7th AA codon through the 3' nucleotide of the 18th AA codon.
 XX Each primer contd. an EcoRI site at its 5' end. The first pair of
 XX oligos are AAQ46882/046883. The second pair are AAQ46884/046885. The
 XX probe to screen an E18 chick brain cDNA library. AAQ46887 is the
 XX insert from a positive clone obtd. by the screening an E18 chick
 XX brain cDNA library in lambda gt10 and introduced into a
 XX plasmid designated p65-21.




```

; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,248
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/801,263
; FILING DATE: 19-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11717 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-102-248-4

alignment_scores:
    Quality: 92.00      Length: 283
    Ratio: 0.724       Gaps: 18
    Percent Similarity: 44.87%   Percent Identity: 24.735

alignment_block:
US-09-674-779-2 x US-09-102-248-4 ..
Align seg 1/1 to: US-09-102-248-4 from: 1 to: 11717

11 ThrLeulleSerSerMetLeuValAlaCysSerAlaProleProThrAs 27
||||| ||| ||||| ||||| ||||| ||||| |||||
8664 ACCTTGACGAGCCCGTACTTTGGGCACAATGCTGTCTACCATACTGA 8713

27 nProGlnValSerProfileLysThrProSerValLeulleThrylAspL 44
:||| ||||| ||||| ||||| ||||| ||||| |||||
8714 ACCGTGCTTTAGCCCGATTAAAGATCGAGCATGCTGG.....GATG 8754.

44 ytleGlyasp..HisHisThrHisGluHisAspGluSerValSerHisV 60
:: ||| ||||| ||||| ||||| ||||| ||||| |||||
8755 AAGCGAGCACCAACACCATACGATACAGACTTCGCCCGCAGTT..... 8797

60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaThr 76
||| ::||| ||||| ||||| ||||| ||||| |||||
8798 .....TGATACGACCAAAGCGGAGCAGC 8821

77 LysGlnGluValVal.ArgTyrClnAlaTyrlLeuGln..... 88
||| ::||| ||||| ||||| ||||| ||||| |||||
8822 AAGCTCAATAAAGTACCGTACATGCTCGCTCGAGCAGGAGTACACCGTCA 8871

88 ..... 88
8872 AAGAGGCACATATGGATGACATCAAGATCAGCACCTCAGGACCGGTAGA 8921

89 .....SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
||| ||| ||||| ||||| ||||| ||||| |||||
8922 AGGCTTAGCTACAAAGGATACCTTCTCCTCGGAAGTGTCTCCAGGGGA 8971

98 ...MetSerGlnLeuLeuThrThraAlaArgSerTrpGlnAlaCys.... 111
::::: ::::::::::::::: ||||| ||||| ||||| |||||
8972 CAGCGTAACGGTTAGTATAGCAGGTAGCAACATCAGAACGTATGCACAA 9021

112 euGlyHisGluProTyrGlnLeu 118
||| ::||| ||||| ||||| ||||| |||||
9022 TGGCCCCCAAGATAAACCAAAATTCGTGGGAGCGGAAAAATATGACCTA 9071

119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
||||| ||| ||||| ::||| ||||| ||||| ||||| |||||
9072 CCTCCCGTTAC.....GATAAGAAGATTCTCTTGACACA...GTGTACGA 9112

135 nAspLeuLys.....SerArgGlylleLeu.....ProAla. 145
: ||||| :||| ||| :|||
9113 CGGTCTGAAGAAGAACACACCCCGGCTACATCATGATGCAGCGCGGAC 9162

146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
::||| ||||| ||||| ||||| ||||| ||||| |||||
9163 GCACGCGCTATACGCTCTCTGTGGAGGAATCATCAG..... 9198

162 agLyCyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIlet 179
||| ::||| ||| ||| |||
9199 .GGAAAGTCTACGCGAGCCACCAT..... 9222

179 rpValProAspLeuGluLeuLysSerGlnAla.LeuTyrGluLeuGlnAs 195
||||| ::||| ||||| ||||| ||||| ||||| |||||
9223 .....CCGGAAGAAGACATATAGTACGAGTGAAGTGGCGGATATACAGA 9267

195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
:||| ||| ||| :|||
9268 CCGG.....TACCGTTACGACCCCGTACCGAAATCAC..... 9298

212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
||||| ::||| ||||| ||||| ||||| ||||| |||||
9299 ...GGGTGACCCGCATCAACGAGTGCCTGCCCTATTAAGACGCGCAACAC 9346

229 LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
::||| ||||| ||||| ||||| ||||| ||||| |||||
9347 GAAGTGGGTCTCAATTCGCCGAGCTTGATCAGACATGCCGA 9388

seq_name: /cgn_2_6/ptodata/2/ina/6A_COMB.seq.US-09-102-248-4
seq_documentation_block:
; Sequence 4, Application US/09102248
; Patent No. 608035
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; NUMBER OF INVENTIONS: Expression of heterologous genes in the bone marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seitzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: NC 280035th Carolina

```



```

112 .....GlyHisGluProTyrGlnLeu 118
1022 TGGCCCGCAAGATAAAACCAAAATTCGTGGACGGGAAATATGACCTA 9071
119 ProProGluHisLeuTyrGlnLeuValProThrLeuHisLeuTyrG1 135
1072 CTCCCGGTCAC.....GTTAAGAAGATTCTCTGGCACA...GTGTACGA 9112
135 nAspLeuLys.....SerArgGlyLeu.....ProAla. 145
9113 CCGCTGAAGAACACACCGCGGCTACATCTATGCACAGGCGGGAC 9162
146 AsnThrGlnLeuSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
9163 CGCAGCGCTATACGTCCTCTCTGAGGAATCATCAG..... 9198
162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIle 179
9199 ..GAAAGTCTACGCGAAGCCACCAT..... 9222
179 rpValProAspLeuGluLeuLysSerGlnAla..LeuTyrGluLeuGlnAs 195
9223 .....CGGAAGAACATAGTACGACGAGTCAAGTGGCGGATACAGA 9267
195 nArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheGlyL 212
9268 CCGG.....TACCGTTACGACCGGTACCGAAATCAC..... 9298
212 euGlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArg 228
9299 ..GGCTGACCGCGCATCAAGACGAGTGGTGGCTATAGAGCAGCACCAAA 9346
229 LysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArg 242
9347 GAAGTGGTCTTCATTCGCGGACTGTATCAGATGCCGA 9388

seq_name: /cqn2.6/ptodata/2/ina/5B_COMB.seq:US-08-910-731-5
seq_documentation_block:
; Sequence 5, Application US/08910731
; Patent No. 5932440
; GENERAL INFORMATION:
; APPLICANT: CHATTERJEE, DEB K.
; APPLICANT: SHANDILYA, HARINI
; TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSER: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., N.W., SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,731
; FILING DATE: (Herewith)
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/795,395
; FILING DATE: 04-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/794,546
; FILING DATE: 03-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/024,057
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: ESMOND, ROBERT W.

```

```

; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942.3440003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; US-08-910-731-5

alignment_scores:
Quality: 86.00 Length: 268
Ratio: 0.694 Gaps: 15
Percent Similarity: 46.269 Percent Identity: 25.373

alignment_block:
US-09-674-779-2 x US-08-910-731-5/rev ..

Align seg 1/1 to reverse of: US-08-910-731-5 from: 1 to: 1386

17 LeuValAlaCysSerAlaProIleProThrAsnProGlnValSerProI1 33
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1101 CTGGCCAGGCGCTGCGACAGCTCCGACGCCGCCATCCTCCAGCCTGT 1052

33 elystrProSerValLeuIleThrLysAspLysIleGlyAspHisHist 50
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1051 TGTTCCTTATCTGTAGCTCCAGGAGAAACCTGTTCTGGGCCAGCAC... 1006

50 hrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe 66
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1005 .....TGAGCTGAAGTGGGGCAGCGCGGCTGT 976

67 GluThrTrpLeuGlnMetHisAlaThrLysGlnGluValValArgTy 83
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
975 GAAGCT...GCAGGAGCTTCAACCCACAGCAGCTCCAGCTGGCAGCAGGTT 929

83 rGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMet..... 98
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
928 CCAGCAGGCTTCCACACAGCTGGGACCCCTCATCCCCAGCTCGTTG 879

99 .....SerGlnLeuLeu.....ThrThrAla 105
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
878 CCGGCCAGGCTGAGCTCCTTCAGGCTCTCTGCGCTGAGGACACACGCA 829

106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProPro..... 120
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
828 CAGATCCCGCAGCCCTTGGCAGTGATGCCACACTCCCATCCAGATCCACAGG 779

121 .....GluHisLeuTrpGlyGlnIleValPro 130
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
778 TCCTGAGCCTGGAGCTGGGGTGAGCAGCCCTGGGACAGCTC...CGCC 732

130 hrLeuHisLeuTyrGlnAspLysSerArgGlyIleLeuProAlaAsn 146
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
731 ATGCCCATCATCACCCAGCTGTTGCTGCCAGGCGCAGCTCCCGCAGCGA 682

147 .....ThrGlnIleArgSerValTyr..... 153
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
681 GCCCTTGGAGCCCAATGCGCACAGTCCCGCAGTGTCTGATGTC 632

154 .....ArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla.... 165
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
631 CACGCGAGCTCTCCAGCTTGGAGCGCTCCAGCTGGCAGGGGAGTCTTC 582

166 ..AlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
581 AGGCCCTGGCACAGCACACGCGACGCCAGCTCATGTATGTCGTTGTGCT 532

```


20827 GGAATACCGAGGSCGGAAGAGGCTTGGAGACTTACCCCATCCACC 20778
213GlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
:::||||| ||| |||||:::|||||:::|
20777 CACCAAGAGCAAAAGCTCTTTGAACAGAGGCTCTTAACCTAAGTGCC 20728
225 lnglyPheArgLysTrpGlyAlaGlnPheSerGluThrAsnSer..... 239
:::||||| |||||:::|||||:::|
20727 ATGGAACAGAT.....GGTGTTCAGGATCCCAAGAGTCTCCGAAATC 20684
240IleCysArgHis.....ValLeuProLys 247
:::||||| ||| |||||:::|||||:::|
20683 GTTGACAGAACTGTGTGCACACTGGCGGTCTCCCTCGG 20642

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-07201-43

seq_documentation_block:
; Sequence 43, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07201
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,841
; FILING DATE: 30-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22481 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: P1-147
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: full length genomic
; OTHER INFORMATION: sequence for PDF plus flanking sequences.
PCT-US95-07201-43

alignment_scores:
Quality: 86.00 Length: 298
Ratio: 0.662 Gaps: 14
Percent Similarity: 43.624 Percent Identity: 21.141
alignment_block:
US-09-674-779-2 x PCT-US95-07201-43/rev ..
Align seg 1/1 to reverse of: PCT-US95-07201-43 from: 1 to: 22481
23 ProileProThrAsnProGlnValSerProIleLysThrProSerVal.. 38
||| |||:::||||| ||| |||||:::|
21479 CCTGCCCCACCCCATCCA.....CCCACCAAGAGCCGAAAGGGTCTT 21439
39LeuIleThrL 42
21438 CAGACATACTCATCTCCTGCAGGACTTGGTGACTTCGCCTTCGTAATC 21389
42 ysAspLysIleGlyAspHisThrHisGluHisAspGluSerValser 58
:::|||||:::|||||:::|
21388 AGCTTCAGCTTGGGACAGTGAGGACCCCTGCACGCTCTCAGTTCTCG 21339
59HisValGlyLeuGlnAlaHisPheGluThrTrpLe 70
:::|||||:::|||||:::|
21338 GTCTATGTCATGATGAATGAACTCGGAGGTGAGCTCTCCTCATCAAGTCA 21289
70 uGlnMethHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrL 87
::: ||| |||||:::|
21288 AATTCTGGGTCACTTTCAGGGGCGAGGAAGATGATACATCATCTCCG 21239
87 euGlnSerArgLeuGlyAsnTyr..... 94
|||:::|||||:::|
21238 GTCAAGGCGAGCTGGGCAATCTGAGACAAGTGAAGCATGTGGTAGT 21189
95 ...LeuProMetSerGlnLeuLeuThrAlaArgSerTrpGlnAl 110
||| |||||:::|
21188 CCTTCAGAGCCCAAGCCGGGGGACAGACAGCCAGGAGC..... 21147
110 aCysGlyHisGluProTyrGlnLeuProGluHisLeuTyrGlyGln. 126
|||:::|||||:::|
21146 TGCCCTTCCCTCTCCGTGACGGGGGCC.....TGCTCGTC 21110
127 ileValProThrLeuHisLeuTyrGlnAspLeuLysSerArgglyIleLe 143
::: |||||:::|||||:::|
21109 CTTATCCAGCCCTCATCTATTTTACAATCCCTTCCAGGTATAGT 21060
143 uProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnG 160
|:::|
21059 TTCTGGCAGAGAGCCA.....TCTCCTGAGGTTCTT 21028
160 lncysAlaGlyGlyAlaAlaMetSer.....Lys 169
::: |||||:::|
21027 CAGAAAGAGTGTGTGGCCCTTTCTTACCTTGGCCCTAATACCACTGAA 20978
170 HisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLy 186
::: |||||:::|
20977 GTTGCTCTGCGTGAAGAATGTTCTCTGCTGTGTTAAGGAGCTAGAAATGAA 20928
186 sSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCys..... 198
: |||||:::|
20927 TATCAGGTCTTATTGTCTCCCAAGTCCATCGGTGTACATTTACTGCTC 20878
199GlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPhe 210
: |||||:::|
20877 ATTTGTTAAATAGAGAGCACTGGACTAAGACAGAGGCGAGAAACCCGGGNA 20828
211 GlyLeu..... 212
|||:::|
20827 GGAATACCGAGGSCGGAAGAGGCTTGGAGACTTACCCCATCCACC 20778
213GlyLeuTyrAlaThrGlyAlaIleHisLeuAspThrG 225
:::||||| ||| |||||:::|||||:::|

```
20777 CACCAAGACAAACGCTCTTTGAACAGAGCTCTTAACCTAAGGTCCC 20728
225 InGlyPheArgLysTrpGlyAlaGlnPheSerGluThrAsnSer..... 239
20727 ATGGAACAGAT.....GGTGTTCAGGATCCCAAGAGTCTCTCGAAATC 20684
240 .....IleCysArgHis.....ValLeuProLys 247
20683 GTTGACAGAAAGCTGTGTGCACACCTCGCGGTGTCTCCCTCGG 20642
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
alignment_scores:
Quality: 86.00 Length: 183
Ratio: 1.075 Gaps: 11
Percent Similarity: 43.716 Percent Identity: 24.044
alignment_block:
US-09-674-779-2 x US-09-103-840A-2/rev ..
Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765
11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProTh 26
4296417ACCTTGACCTCCCGATCGGTGTCCTCGGCATAGCACCGAGCTCACCAT 4296368
26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
4296367CAACAGAGCTGATCGATCCGCGCATGTGTCGCTCG..... 4296334
43 sPLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59
4296333..GAAATCCGCGATCGTATTTCGCG.....ATCGACCAC 4296301
60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaTh 76
4296300CGCGCG.....AGCTATCTGCAGATCATTTTCGCT 4296269
76 rLysGlnGlu.....ValValArgTyrGlnAlaTyr..... 86
4296268CGCCACCGCCCTTCGCGGCGCTACAGGCACTCAACGATCCGA 4296219
87 ..LeuGlnSerArgLysGlyAsnTyrLeuProMetSerGlnLeu 102
4296218CATGCGGATCATGCGGATTTCTGTACACCGAGCAG..... 4296178
103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
```

```
4296177...GTTACAGCAGTGGGAGGATTGC..... 4296154
119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
4296154..... 4296154
136 sPLysSerArgGlyIleLeuProAlaAsn.....ThrGln 148
4296153.....CGCGGTGGATCGTCCCGCGCGGTGGTGGTTCGAG 4296112
149 IleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAl 165
4296111ATCCCGTCTGCTGCAT...GATCCAGCCCTGGCCCGCGGGTAAGCAGCG 4296065
165 aAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValPro 181
4296064CGCG.....TCGGGCTTCGCAATGTGTTCCCG 4296037
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
alignment_scores:
Quality: 86.00 Length: 183
Ratio: 1.075 Gaps: 11
Percent Similarity: 43.716 Percent Identity: 24.044
alignment_block:
US-09-674-779-2 x US-09-103-840A-1/rev ..
Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529
11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaPro...IleProTh 26
4304162ACCTTGACCTCCCGATCGGTGTCCTCGGCATAGCACCGAGTCCACCAT 4304113
26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
4304112CAACAGAGCTGATCGATCCGCGCATGTGTCGCTCG..... 4304079
43 sPLysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHis 59
4304078..GAAATCCGCGATCGTATTTCGCG.....ATCGACCAC 4304046
60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaTh 76
4304045CGCGCG.....AGCTATCTGCAGATCATTTTCGCT 4304014
76 rLysGlnGlu.....ValValArgTyrGlnAlaTyr..... 86
4304013CGCGAGCAGCCGCGCTTCGCGGCGCTTACAGGCACTCAACGATCCGA 4303964
```



```

alignment_scores:
  Quality: 85.00      Length: 59
  Ratio: 2.179       Gaps: 0
  Percent Similarity: 66.102  Percent Identity: 30.508

alignment_block:
  US-09-674-779-2 x US-08-673-312-9/rev ..

  Align seg 1/1 to reverse of: US-08-673-312-9 from: 1 to: 61662

32  proilelythrProSerValLeulleThrLysAspLysleGlyAspHi 48
    ||| :||| :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2766 CCATATTCAACGGGAAACGCTTGCTCGAGGGTGACAAGGCAGGAACCA 2717
    48 sHsIthrHisGluHisaspLuserValSerHisValGlyLeuGlnAlaH 65
    ||| |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2716 CCAATCAACACGACCGCGCAACAAAGAGCGGGCGCTGCCCCCATC 2667
    65 isPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluVal 81
    :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2666 AAATTCAGACATAATCATCAATCGCGTGAAGGCAGCACGTTTAC 2617
    82 ArgTyrGlnAlaTyrLeuGlnSerArg 90
    :: |||| |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
2616 GCCATTACGCCTACCCACATTCCTCCGC 2590

seq_name: /cqn2_6/ptodata/2/ina/5A_COMB.seq:US-08-276-099A-1

seq_documentation_block:
; Sequence 1, Application US/08276099A
; Patent No. 5591825
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L
; APPLICANT: Hou, Jinhao
; TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND

```

```

alignment_scores:
  Quality: 83.50      Length: 118
  Ratio: 1.392      Gaps: 7
  Percent Similarity: 50.847      Percent Identity: 29.661

alignment_block:
  US-09-674-779-2 x US-08-276-099A-1/rev ..

  Align seg 1/1 to reverse of: US-08-276-099A-1 from: 1 to: 3046

66 PheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnLuValValAr 82
   |||:::|||||
1016 TTCAGTACCTGGGGGGCTCTTCACCAAGCAAGCAACTGGTCAGGAG 967
   |||:::|||||

82 gTyGlnAlaTyLeuGlnSerArgLeuGlyAsnTyLeuProMetLS 99
   |||:::|||||
966 GGTTCTCAGGACTTCATCCACCGG.....CCAGTCA 935
   |||:::|||||

99 erGlnLeuLeuThrThrAlaArg, SerTrpGlnAlaCysGlyHisGluPr 115
   |||:::|||||
934 GCGAT.....GCCCGGCTTGGGTCAAGTCCCCACGAGCC 897
   |||:::|||||

115 oTyGlnLeuProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
   |||:::|||||
896 G....CCCTACCTCCTGCTAGCTGGGATAAATGTCACCAGGCGTTTC 850
   |||:::|||||

132 isLeuTyGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGln 148
   |||:::|||||
849 AC.....ACCTCTCCTGGAGTGGGCGCAGGCTCTCCTCAA 815
   |||:::|||||

149 ....IleArgSerValTyArgAsnProGluLeuAsnGlnCysAlaGly 164
   |||:::|||||
914 AACGACGACG.....CATTCCTCGCAGCTGC 789
   |||:::|||||

```

```

164 yAlaA1aMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValp 181
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||
788 TGTGCGGGTTCACAAATCTGGATCCTCTTCAGCACTAGGCTTTGGCTGC 739

181 ro 181
       ||
738 CT 737

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-781-890-1

```

seq_documentation_block:
; Sequence 1, Application US/08781890
; Patent No. 5710266
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L
; APPLICANT: Hou, Jinchao
; TITLE OF INVENTION: INTERLEUKIN-4 SIGNAL TRANSDUCERS AND
; TITLE OF INVENTION: BINDING ASSAYS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSER: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; Zip: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781.890
; FILING DATE: 05-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/276,099
; FILING DATE: 15-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59451-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3046 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 166..2706
; OS-08-781-890-1

```

alignment_scores:		
Quality:	83.50	Length: 118
Ratio:	1.392	Gaps: 7
Percent Similarity:	50.847	Percent Identity: 29.661

alignment_block:

US-09-674-779-2 x US-08-781-890-1/rev

Align seg 1/1 to reverse of: US-08-781-890-1 from: 1 to: 3046

66 PheGluThrTrpLeuGlnMethHisHisAlaThrLysGlnGluValVal 82
 |||||
 1016 TTCAGTACCTGGGGGGTCTTCCACCAGGAGCACTGGTGACGAG 967

```

82  gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPromets 99
    |
    |:::|||||
966  GGTTCACGACTTCATCCAGCCGG.....CCAGTCA 935
    |
    |:::|||||
99  erGlnLeuLeuThrThrAlaArg_SerTrpGlnAlaCysGlyHisGluPr 115
    |:::|
    |:::|||||
934  CGCAT.....GCCCGGTCTTGGGCTCAAGCTCCCCACGACGC 897
    |
    |:::|||||
115  oTyrGlnLeuProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
    |
    |:::|||||
896  G....CCCTACCTCTGCTAGCTGGGAATAATGCCACGAGCTTTC 850
    |
    |:::|||||
132  isLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaasnThrGln 148
    ||
    ||:::|||||
849  AC.....ACCTCTCCTGGAGTGGGGCAGGCTCTCCTCAA 815
    |
    |:::|||||
149  ...IleArgSerValTyrArgAsnProGluLeuasnGlnCysAlaGlyCl 164
    |
    |:::|||||
814  ACGGTGGC.....CATTCCTGCCAGCTGC 789
    |
    |:::|||||
164  yAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValp 181
    |:::|||||
    |:::|||||
788  TGTGCGGTTTCCAAATCTGGATCCTCTTCAGCACTAGGGCTTTGGCTGC 739
    |
    |:::|||||
181  ro 181
    ||
738  CT 737

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-087-465-11

```

seq_documentation_block:
; Sequence 11, Application US/09087465A
; Patent No. 6160092
; GENERAL INFORMATION:
; APPLICANT: Vinkemeier, Uwe
; APPLICANT: Chen, Xiaomin
; APPLICANT: Darnell Jr., James E
; APPLICANT: Kuruyan, John
; TITLE OF INVENTION: A CRYSTAL OF THE CORE
; TITLE OF INVENTION: USE
; FILE REFERENCE: 600-1-229
; CURRENT APPLICATION NUMBER: US/09/087,465A
; CURRENT FILING DATE: 1998-05-29
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3046
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-087-465-11

```

alignment_scores:		
Quality:	83.50	Length: 118
Ratio:	1.392	Gaps: 7
Percent Similarity:	50.847	Percent Identity: 29.661

alignment_block:

US-09-674-779-2 x US-09-087-465-11/rev

Align seg 1/1 to reverse of: US-09-087-465-11 from: 1 to: 3046

66 PheGluThrTrpLeuGlnMethHisHisAlaThrLysGlnGluValVal 82
|||:::||||| :: ||||| |||:::|
1016 TTCAGTACCTGGGGGGGGTGTCTTCCACCAGGAAGCAACTGGTGACGAG 967

82 gTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPmets 99
966 GTTCTCAGGAGTTCATCCAGCGG.....CCAGTCA 935

99 erGlnLeuLeuThrAlaArg_SerTrpGlnAlaCysGlyHisIslupr 115
934 GCGAT.....GCCCGGGTCTTGGGCTCAAGTCCGCCACAGCG 897

seq name: /cgn2 6/ptodata/2/ina/5A COMB, seq: US-07-621-193A-1

```

SEQUENCE CHARACTERISTICS:
LENGTH: 4527 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Herpesvirus of turkey
STRAIN: PB-THV1
FEATURE:
NAME/KEY: CDS
LOCATION: 1..81
OTHER INFORMATION: /label= end_of_
FEATURE:

```

```

; NAME/KEY: CDS
; LOCATION: 316..945
; OTHER INFORMATION: /label= ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1084..2124)
; OTHER INFORMATION: /label= ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2322..3170)
; OTHER INFORMATION: /label= ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3320..4504
; OTHER INFORMATION: /label= ORF5
;
; US-07-621-193A-1
;
alignment_scores:
    Quality: 82.50      Length: 201
    Ratio: 0.743      Gaps: 13
    Percent Similarity: 55.224      Percent Identity: 24.378
;
alignment_block:
US-09-674-779-2 x US-07-621-193A-1 ..
Align seg 1/1 to: US-07-621-193A-1 from: 1 to: 4527

14 SerSerMetLeuValAlaAlaCysSerAlaProIleProThrAsnProGlnVa 30
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2866 TCAACAGCTCCACAAATAATGTCATCTCTCGTCTCTGCCAATCCGAACC 2915
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
30 lser...ProIleLysThrProSerValLeuIleThr..... 41
      ::|||:::|||||:|||||:|||||:|||||:|||||:|||||:
2916 GGGCATACCACCTCCACACCTGCCGATTTAATTCACAATTGGCGCATGCC 2965
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 ..LysAspLysIleGlyAspHisHisThrHisGluHisAspGluSerVal 57
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2966 GCGCGGGCAAAACCAATGTGGATTGGCAACCGCACAGGTCCTCTGTA 3015
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 ..SerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethI 73
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3016 CGGACTAATATGGGCACACCCACACATCATTC.....TTCAGATGCTCCAT 3059
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 sHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerA 90
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3060 GCATTGTCTATGAGAAAGATCCATAGGGTGAGGCGCGTCACCGAGATC 3109
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
90 rgLeuGlyAsnTyrLeuProMetSerGlnLeuLeuThrThrAlaArg 106
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3110 GCCCAGGCAAT.....CGATCGCATTCGTCTAGTAAA 3141
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 SerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProProGlu..... 121
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3142 GTGACGAGAGTTATCATGCACACACCCATGCCCGCTTCGGAATAACT 3191
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 ..HisLeuTrp.....GlyGlnIleValProThrLeuHisLeuTyrGlnA 136
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3192 GGAGCTGTGGAGATCGGAAGCTCTTTTGACTGCCGGTCTCGTA.... 3237
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 spLeuLysSerArg..GlyIleLeuProAlaAsnThrGlnIleArgSerVa 152
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3238 ..CTACTTTCGCACAGGTGTATACCGGACGGCTACTATATTTTATAT 3285
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 lTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerL 169
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3286 CATTCACAGCTCCGAAATTTACATACGTCG.....GCGGCGATGGAAG 3326
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 ysHisLeuThrAsnSerAlaIle.....Asp..IleTrpValProAspLe 183
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3327 TAGATGTTGAGTCTTCGAAGTAAGTGCTCGAATATGGGTATGTCTGT 3376
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 uGluLeuLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGln 199

```



```

; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,214
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1110-139P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: mouse
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..837
; US-08-339-214-23

```

alignment_scores:

```

Quality: 82.00 Length: 253
Ratio: 0.837 Gaps: 16
Percent Similarity: 38.735 Percent Identity: 25.296

```

alignment_block:

```

US-09-674-779-2 x US-08-339-214-23 ..
Align seg 1/1 to: US-08-339-214-23 from: 1 to: 837
23 profileProthrAnProGlnValSerProIleLysThrProSer..... 37
133 CCGCCACTCCACCACCTGTCACCTACCCCGCCATCACAAACC 182
38 .....ValLeuIleThrLysAspLysIleGlyAspHisThrH 51
183 ACTCCCACTGCGGCCTGACCCCTCTAAAGAGAGAGACCAACAA 232
51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
233 AT..... 234
68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnValValArgTrpG1 84
235 CTGTGGCTA.....CCGGTGGTATTTCAT 260
84 n.....AlaTrpLeuGlnSerArgLeuGlyAsnTrpLeuProPom 98
261 GGTCTGTGGTCTGCTGTTGGAATGGGATAGGAATGAT..... 300
98 etSerGlnLeuLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisG1 114
301 .....CAGCTCTCCACTGCAGAGAACTGGCAGAACTCCCGTGGCTTC 345
114 uProTrpGlnLeuProProGluHisLeuTrpGlyGlnIleValProthr. 130
346 ACCAACCAAGCCCTTAAGATGATCATCTTTTGAAGAGCAATAGCCCAACC 395

```

```

131 ..LeuHisLeuTrpGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
396 CAGTACACCCCTCTGAAAAA...AAGAGCCGAGAGGTGTGCCCATTTAA 442
146 nThrGlnIleArgSerValTrpArgAsnProGluLeuAsnGlnCysAlaG 163
443 C.....AGGGAACCCCACTCAAGGTCCATCCCT 471
163 lGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
472 CTGGA.....ATGGGAAGACATATGGAAACCCCTCT...GATCTCTGG 512
179 ..... 179
513 AGTGAAGTATAAGAAAGGTGGCTTGTGATCAACGAAACTGGGTGTACT 562
180 .....ValProAspLeuGluIle..... 185
563 TCGTGTATTCAAAAGTATACCTCCGGGTGAGTCTTGCAACAAACCAAGCC 612
186 ..LysSerGlnAlaLeuTrpGlu..... 192
613 CTAACCAACAAGGTCTATATGAGGAAGTCTTAAGTATCTGAGGATCTGGT 662
193 .....LeuGlnAsnArgLeuCysGlnTrpTrpLeuGluHisG1 205
663 GCTAATGAGGAGAGAGAGGTTGAAGTACTGCACTACTGGACAGATATGGG 712
205 yGluAsnGln.AsnPheGlyLeuGlyLeuTrpAlaThrGlyAlaIleHis 221
713 CCCAGCAGCAGCTACTGGGGCAGTATTCATCTTACCAGTCTGACCAT 762
222 Leu 222
763 TTA 765
seq_name: /cgn2_6/ptodata/2/ina/68_COMB.seq:us-08-339-214-36
seq_documentation_block:
; Sequence 36, Application US/08339214
; Patent No. 6348334
; GENERAL INFORMATION:
; APPLICANT: Nagata, Shigikazu
; APPLICANT: Suda, Takashi
; APPLICANT: Takahashi, Tomoniro
; APPLICANT: Nakamura, No. 6348334io
; TITLE OF INVENTION: A Fas Ligand, A Fragment Thereof and DNA
; TITLE OF INVENTION: Encoding the Same
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/339,214
; APPLICATION NUMBER: 28,977
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1110-139P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050

```


795 TTA 797

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-339-214-31
seq_documentation_block:
; Sequence 31, Application US/08339214
; Patent No. 6348334
; GENERAL INFORMATION:
; APPLICANT: Nagata, Shigikazu
; APPLICANT: Suda, Takashi
; APPLICANT: Takahashi, Tomonori
; APPLICANT: Nakamura, No. 6348334io
; TITLE OF INVENTION: A Fas Ligand, A Fragment Thereof and DNA
; TITLE OF INVENTION: Encoding the Same
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,214
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1110-139p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: mouse
; IMMEDIATE SOURCE:
; CLONE: LambdaMFL5, LambdaMFL18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 125..961
; OTHER INFORMATION: /label= Figs_23-24
US-08-339-214-31

alignment_scores:
Quality: 82.00 Length: 253
Ratio: 0.837 Gaps: 16
Percent Similarity: 38.735 Percent Identity: 25.296

alignment_block:
US-09-674-779-2 x US-08-339-214-31 ..
Align seg 1/1 to: US-08-339-214-103 from: 1 to: 927
23 profileProthrAsnProGlnValSerProIleLysThrProSer..... 37
165 CCGCCACCTCCACCACCACTGCTGTCACCTACCCACCGCCATCAACAC 214
38ValLeuIleThrLysAspLysLleGlyAspHisHisThrH 51
215 ACTCCCACTGCCGCACTGACCCCTCTAAAGAGAGAGGACCAACACAA 264
51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
265 AT..... 266
68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
267 CTGTGGCTA.....CCGGTGGTATTTCAT 292
84 n.....AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProm 98
293 GGTTCGTGGCTCTGTTGGAATGGATTAGGAATGAT..... 332
98 etSerGlnLeuLeu,ThrThraArgSerTrpGlnAlaCysGlyHisGl 114
333CAGCTCTCCCTCGCAGAGGAACTGGCAGAACTCCGTCAGTTC 377
114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
378 ACCAACCAAGCGCTTAAAGTATCATCTTTGAAAGCAATAGCAACCC 427
131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
428 CAGTACACCTCTGAAAAA...AAGACCGGAGAGTGTGGCCCATTTAA 474
146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
475 C.....AGGAACCCCACTCAAGGTCCATCCCT 503
163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
504 CTGGA.....ATGGGAGACACATATGAACCGCTCT...GATCTCTGG 544
179 179
545 AGTGAAGTATAAGAAAGTGGCCTGTGTATCAACGAAGTGGTGTACT 594
180ValProAspLeuGluIle..... 185
595 TCGTGATTCCAAAGTATCTCCGGGGTCAGTCTTGCAACAAACAGGCC 644
186 ..LysSerGlnAlaLeuTyrGlu..... 192
645 CTAACACCAAGGTCTATATGAGAACTTAAGTATCTCTGAGGATCTGTT 744
193LeuGlnAsnArgLysCysGlnTyrTrpLeuGluHisGl 205
695 GCTAATGGAGGAGAGAGAGTGTGAACCTACTGCTGACGACAGATATGG 744
205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
745 CCCACAGCAGCTACCTGGGGGAGTAGTATCAATCTTACCAGTGTGCTAC 794
222 Leu 222
|||

alignment_scores:
Quality: 82.00 Length: 253
Ratio: 0.837 Gaps: 16
Percent Similarity: 38.735 Percent Identity: 25.296

alignment_block:
US-09-674-779-2 x US-08-339-214-103 ..
Align seg 1/1 to: US-08-339-214-103 from: 1 to: 927
23 profileProthrAsnProGlnValSerProIleLysThrProSer..... 37
165 CCGCCACCTCCACCACCACTGCTGTCACCTACCCACCGCCATCAACAC 214
38ValLeuIleThrLysAspLysLleGlyAspHisHisThrH 51
215 ACTCCCACTGCCGCACTGACCCCTCTAAAGAGAGAGGACCAACACAA 264
51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
265 AT..... 266
68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGl 84
267 CTGTGGCTA.....CCGGTGGTATTTCAT 292
84 n.....AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProm 98
293 GGTTCGTGGCTCTGTTGGAATGGATTAGGAATGAT..... 332
98 etSerGlnLeuLeu,ThrThraArgSerTrpGlnAlaCysGlyHisGl 114
333CAGCTCTCCCTCGCAGAGGAACTGGCAGAACTCCGTCAGTTC 377
114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
378 ACCAACCAAGCGCTTAAAGTATCATCTTTGAAAGCAATAGCAACCC 427
131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
428 CAGTACACCTCTGAAAAA...AAGACCGGAGAGTGTGGCCCATTTAA 474
146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
475 C.....AGGAACCCCACTCAAGGTCCATCCCT 503
163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
504 CTGGA.....ATGGGAGACACATATGAACCGCTCT...GATCTCTGG 544
179 179
545 AGTGAAGTATAAGAAAGTGGCCTGTGTATCAACGAAGTGGTGTACT 594
180ValProAspLeuGluIle..... 185
595 TCGTGATTCCAAAGTATCTCCGGGGTCAGTCTTGCAACAAACAGGCC 644
186 ..LysSerGlnAlaLeuTyrGlu..... 192
645 CTAACACCAAGGTCTATATGAGAACTTAAGTATCTCTGAGGATCTGTT 744
193LeuGlnAsnArgLysCysGlnTyrTrpLeuGluHisGl 205
695 GCTAATGGAGGAGAGAGTGTGAACCTACTGCTGACGACAGATATGG 744
205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
745 CCCACAGCAGCTACCTGGGGGAGTAGTATCAATCTTACCAGTGTGCTAC 794
222 Leu 222
|||

```

307 ACTCCACTGCCGCTGACCTCTAAAGAAAGAGGACCAACACAA 356
51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
357 AT.....
68 ThrTrpLeuGlnMetHisAlaThrLysGlnGluValValArgTyrGln 84
359 CTGTGGCTA.....CCGGTGGTATTTCAT 384
84 n.....AlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPom 98
385 GGTTCGTGCTGCTGCTGGAATGGATAGGAATGAT.....424
98 etSerGlnLeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGln 114
425 ....CAGCTCTTCCACCTGTCAGAGGAAGTGGCAGAACTCGTGATTC 469
114 uProTyrGlnLeuProProGlnHisLeuTrpGlyGlnIleValProThr. 130
470 ACCAACCAAGCCTTAAAGTATCATCTTTTGAAGCAAAATAGCCAAACC 519
131 ..LeuHisLeuTyrGlnAspLeuSerArgGlyIleLeuProAla.As 146
520 CAGTACACCTCTGAAAAA...AAGAGCGAGGAGTGGCCCATTTAA 566
146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
567 C.....AGGAACCCCACTCAAGGTCCATCCCT 595
163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
596 CTGGA.....ATGGCAAGACACATATGGAACCGCTCT...GATCTCTGG 636
179 .....
637 AGTGAAGTATAAGAAAGGTGGCTTGTGATCAACGAAACTGGGTGTACT 686
180 .....ValProAspLeuGluIle.....185
687 TCGTGTATTCCAAAGPATACCTCCGGGTACGTCTTGCAACACAGCC 736
186 ..LysSerGlnAlaLeuTyrGlu.....192
737 CTAACCAACCAAGTCTATATGAGAACTCTAAGTATCTCTGAGGATCTGT 786
193 .....LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGln 205
787 GCTAATGGAGGAGAGAGGTTGAAGTACTGCACTACTGACAGATATGG 836
205 yGluAsnGln.AsnPheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHis 221
837 CCACACAGCACTACCTGGGGGAGTATTTCAATCTTACCAGTCTGACCAT 886
222 Leu 222
887 TTA 889

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-038-832-3

```

seq_documentation_block:
; Sequence 3, Application US/09038832
; Patent No. 6146945
; GENERAL INFORMATION:
; APPLICANT: KIKLY, KRISTINE
; APPLICANT: ERICKSON-MILLER, CONNIE
; TITLE OF INVENTION: Sialoadhesin Family Member-2
; TITLE OF INVENTION: (SAF-2)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980

```

```

CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/038,832
FILING DATE: 11-MAR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/041,886
FILING DATE: 02-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-50018
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1597 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-038-832-3

alignment_scores:
Quality: 81.50 Length: 229
Ratio: 0.849 Gaps: 9
Percent Similarity: 41.921 Percent Identity: 22.271

alignment_block:
US-09-674-779-2 x US-09-038-832-3 ..
Align seg 1/1 to: US-09-038-832-3 from: 1 to: 1597
23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
677 CCTCATCTAGGACCTAGACTGTGCCACTCCAGGAACCTGACCTGCT 726
39 uIleThrLysAspLys...IleGlyAspHisHisThrHisGluHisAspG 55
727 CTGTGCTGCGCTGTAAGCAGGGAC.....ACCCCCCATGAT. 766
55 LuSerValSerHisValGlyLeuGlnAla.....HisPheGlu 67
767 ....CTCCTGTGATTGGGCTCCGTGCTCCCGGGCCCTACTACTGC 811
68 ThrTrpLeuGlnMetHisAlaThrLysGlnGluValValArgTyrGln 84
812 CCCTCTCTCAGTGTCTACCTTACCCCAAGCCAGGACCCAGGACCA 861
84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGlnL 101
862 GCCTACCTGTCAGGTGACCTT.....883
101 euLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
884 .....GCCTGGGACAGGTGTGACCACGACCACTACCGT 916
118 LeuProProGlnHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
917 ...CCGCTCGA.....TGTGCTCTACCTCTCTTGAACCTGA 951
134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151

```

```

952 CCATGAC..... 958
151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyAlaAlaMet 167
959 .....TGTCTTCCAAGAGATGCCAC 979
168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG1 184
980 AGATCCACAGCCCTGGGAATGGCTCATCTCTTCAGTCTCTGAGGGCC 1029
184 uLeLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrT 201
1030 AGTCTCTGGCGCTGGTCTGTGTGTCAACAGCAATCCCTGCCAGGCTG 1079
201 rLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu..... 214
1080 AGCTGGACCGGGGAGCTGACCTGTGCCCTCAGCGTCTCTCAAAACC 1129
215 TyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpG1 231
1130 TGGCTGTCTGGAGCTGCCCTGAGTGCAGGTCCAGGCTCCAGCACAT 1179
231 yAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
1180 CTGCGGAGCTCAGAACGCTCAGGCTCCAGCACAT 1216

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-038-832-1

seq_documentation_block:
; Sequence 1, Application US/09038832
; Patent No. 6146845
; GENERAL INFORMATION:
; APPLICANT: KIKLY, KRISTINE
; APPLICANT: ERICKSON-MILLER, CONNIE
; TITLE OF INVENTION: Sialoadhesin Family Member-2
; TITLE OF INVENTION: (SAF-2)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/038,832
; FILING DATE: 11-MAR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/041,886
; FILING DATE: 02-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: GH-50018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2900 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA

us-09-038-832-1

alignment_scores:
Quality: 81.50 Length: 229
Ratio: 0.849 Gaps: 9
Percent Similarity: 41.921 Percent Identity: 22.271

alignment_block:
US-09-674-779-2 x US-09-038-832-1 ..
Align seg 1/1 to: US-09-038-832-1 from: 1 to: 2900

23 ProIleProThrAsnProGlnValSerProIleLysThrProSerValLe 39
537 CCTCATCTAGGACCTAGAGTCTGCCACTCCAGGAACCTGACCTGCT 586
39 uileThrLysAspLys...ileGlyAspHisHisThrHisGluHisAsp 55
587 CTGTGCCCTGGGCTGTAGCAGGGGAC.....ACCCCCCATGAT. 626
55 luserValSerHisValGlyLeuGlnAla.....HisPheGlu 67
627 .....CTCCTGGATGGGGCTCCGTCTCTCCCGGGGCCCACTACTGC 671
68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG1 84
672 CCGCTCCTCAGTGTACCCCTTACCCCAAGACCCAGGACCCAGGCACCA 721
84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGlnL 101
722 GCCTCACCCTGTAGGTGACCTT..... 743
101 euLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
744 .....GCCTGGACAGGTGTGACCAACGACCACTACCGT 776
118 LeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
777 ....CCGCTCGA.....TGTCTCTACCTCCCTTGGAACTGA 811
134 rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
812 CCATGAC..... 818
151 erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyAlaAlaMet 167
819 .....TGTCTTCCAAGAGATGCCAC 839
168 SerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG1 184
840 AGCATCCACAGCCCTGGGAATGGCTCATCTCTTCAGTCTTTCAGGGCC 889
184 uileLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyrT 201
890 AGTCTCTGGCGCTGGTCTGTGTGTCAACAGCAATCCCTGCCAGGCTG 939
201 rLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeu..... 214
940 AGCTGGACCGGGGAGCTGACCTGTGCCCTCAGCGTCCCTCAAAACC 989
215 TyrAlaThrGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpG1 231
990 TGGCTGTCTGGAGCTGCCCTGAGTGCAGGTGAGGATGAAGGGGAATCA 1039
231 yAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
1040 CTGCGGAGCTCAGAACGCTCAGGCTCCAGCACAT 1076

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-439-313-474

seq_documentation_block:
; Sequence 474, Application US/09439313
; Patent No. 6329505
```

```

; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 474
; LENGTH: 1594
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-439-313-474

```

```

alignment_scores:
  Quality: 81.00      Length: 209
  Ratio: 0.871       Gaps: 10
  Percent Similarity: 44.498      Percent Identity: 20.096

```

```

alignment_block:
  US-09-674-779-2 x US-09-439-313-474
  Align seg 1/1 to: US-09-439-313-474 from: 1 to: 1594

```

```

48 HisHisThrHisGluHisAspGluSerVal.....SerHisValGlyLe 62
|||||
919 CATCACATCATGAACACACACACACATGCCATATGGACACTGGCACAC 968
62 uGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnG 79
|||||
969 ACACCTGCCACACTGTACATGGACACACCTCCACACCATCACACACACC 1018
79 luValValArgTyTGlnAlaTyTGlnSerArgLeuGlyAsnTyLeu 95
|||||
1019 ACACACTGCCT.....GTGCACACAGGACACACACACTGT 1056
96 ProPrometSerGlnLeuThrThrAlaArgSerTrpGlnAlaCysG 112
|||||
1057 CACACAGATACACAAACACACTGTGCACACGGAGA..... 1089
112 yHisGluProTyTGlnLeuProGluHisLeuTrpGlyGlnIleValP 129
|||||
1090 .CATCACCATGCAGATACACACCA.....CTCTGGTGCCTG..... 1125
129 roThrLeuHisLeuTyTGlnAspLeuLysSerArgGlyIleLeuProAla 145
1125 ..... 1125
146 AsnThrGlnIleArgSerValTyArgAsnProGluLeuAsnGlnCysAl 162
|||||
1126 .....CTGAATTACCCCTGC 1139
162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIle 179
|||||
1140 TGGGGGGACAGCAGTGGCA.....TACTCATGCTTAAGTGACT 1177
179 rpVal.....ProAsp 182
|||||
1178 GCCTTTTACCCCATAGTAGTGATGTCCTCCATCAACACTGCCACCCACG 1227
183 LeuGluIleLysSerGlnAlaLeuTyTGlnLeuGlnAsnArgLeu..... 197

```

```

1228 TTGGGGCTA...CCCGACCCCATCTTTTACAAACACAGGCGCAAGTGAACATA 1274
198 .CysGlnTyTrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyL 214
1275 ATGGAGTGGTGGAGGAGTGGAGAAATCCCGCGGTAGTCACCGGGAT 1324
214 euTyAlaThrGlyAlaIleHisLeuAspThrGlnGlnPheArgLysTrp 230
1325 AGAATTCCCAAGGACCCCTCTTTTGGAG...GATGGTTTCCATTTCTGG 1371
231 GlyAlaGlnPheSerGluThrAsnSer 239
1372 AGCGATCTGCCGACAGGCGTGAATGCC 1398
seq_name: /cgn2_5/ptodata/2/1na/5A_COMB.seq:US-08-741-881-1
seq_documentation_block:
; Sequence 1, Application US/08741881
; Patent No. 5789245
; GENERAL INFORMATION:
; APPLICANT: Dubensky Jr, Thomas W
; APPLICANT: Polo, John M.
; APPLICANT: Ibanez, Carlos E.
; APPLICANT: Chang, Stephen M.W.
; APPLICANT: Jolly, Douglas J.
; APPLICANT: Driver, David A.
; APPLICANT: Belli, Barbara A.
; TITLE OF INVENTION: EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/741,881
; FILING DATE: 30-Oct-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 930049.423C6 / 1146.007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-741-881-1

```

```

alignment_scores:
  Quality: 81.00      Length: 210
  Ratio: 0.853       Gaps: 14
  Percent Similarity: 45.238      Percent Identity: 25.238

```

```

alignment_block:
  US-09-674-779-2 x US-08-741-881-1
  Align seg 1/1 to: US-08-741-881-1 from: 1 to: 16656

```

```

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,158
; FILING DATE: 30-OCT-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 930049.423D3 / 1146.012
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-739-158-1
;
;
; alignment_scores:
; Quality: 81.00 Length: 210
; Ratio: 0.853 Gaps: 14
; Percent Similarity: 45.238 Percent Identity: 25.238
;
; alignment_block:
; US-09-674-779-2 x US-08-739-158-1 ..
;
; Align seg 1/1 to: US-08-739-158-1 from: 1 to: 16656
;
; 11 ThrLeuileSerSerMetLeuValalaCysSerAlaProIleProThrAs 27
; ||||| ||| :|||
; 8649 ACCCTGACGAGCCCTACTTGGGCACATGCTGCTACTGCCACCATCTGA 8698
;
; 27 nProGlnValSerProIleuThrProSerValLeuIleThrLysAspL 44
; ||||| ||| :|||
; 8699 ACCGTGCTTACGCGCTGTTAAGATCAGCAGGTCTGG.....GAGC 8739
;
; 44 ysIleGlyasp..HisHisThrHisGluHisAspGluSerValSerHisv 60
; :||| ||||| |||||
; 8740 AAGCGAGCATACACATACGATACAGACTTCCGCCAGTT..... 8782
;
; 60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaThr 76
; ||||| ||| :|||
; 8783 .....TGATACGACCAAGCGGACGACC 8806
;
; 77 LysGlnGluValValArgTyrGlnAlaTyrLeuGln..... 88
; ||||| ||| :|||
; 8807 AAGCGCAACAAGTACGCTGATGCTTAAAGCAGGATCACACCGTTA 8856
;
; 88 ..... 88
;
; 8857 AAGAGCCACCATGGATGACATCAAGATTAGCACCTCAGACCGGTGTAGA 8906
;
; 89 .....SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
; ||||| ||| :|||
; 8907 AGGCTTAGCTACAAAGGATCTTTCTCTCGCAAAATGCCCTCCAGGGA 8956
;
; 98 ....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
; :||| ||||| |||||
; 8957 CAGCGTAACGGTTAGCATAGTGTAGTAGCAACTCAGCAACGTCATGTACAC 9006
;
; 112 .....GlyHisGluProTyrGlnLeu 118
; ||||| ||| :|||
; 9007 TGGCCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056
;
; 119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
; ||||| ||| :|||
; 9057 CCTCCCGTTTCAC.....GGTAAAGAAATTCCTTGCACA...GTGTACGA 9097
;
; 135 nAspLeuLys...SerArgGlyIleLeu.....ProAla.Asn 146
; ||||| ||| :|||
; 9098 CCGTCTGAACAACACTCCAGGCTACATCATCTATGCACGCGCGGACCGC 9147
;
; 147 ThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGl 163
; ||||| ||| :|||
; 9148 ACGCTTATATCTCTACCTGGAAGAATCATCAG.....GG 9182
;
; 163 yGlyAlaAlaMetSerLysHisLeu 171
; ||||| ||| :|||
; 9183 AAAGTTTACGCAAGCGCCGATCTG 9207
;
; seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq;us-08-739-158-1
;
; seq documentation_block:
; Sequence 1, Application US/08739158
; Patent No. 5814482
; GENERAL INFORMATION:
; APPLICANT: Dubensky Jr, Thomas W
; APPLICANT: Polo, John M.
; APPLICANT: Jolly, Douglas J.
; APPLICANT: Driver, David A.
; TITLE OF INVENTION: EUKARYOTIC LAYERED VECTOR INITIATION SYSTEMS
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,796
; FILING DATE: 15-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mcmasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 930049.42305 / 1146.006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16656 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-404-796-1

alignment_scores:
    Quality: 81.00      Length: 210
    Ratio: 0.853        Gaps: 14
Percent Similarity: 45.238      Percent Identity: 25.238

alignment_block:
US-09-674-779-2 x US-08-404-796-1 ..

Align seg 1/1 to: US-08-404-796-1 from: 1 to: 16656

11 ThrLeuIleSerSerMetLeuValAlaLacysSerAlaProIleProThrAs 27
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8649 ACCCTGACGAGCCCTACTTGGGCACATGCTGCTACTGCCACCATACTGA 8698

27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8699 ACCGTGCTTCAGCCCTGTTTAAGATCGAGCAGCTCTGG.....GACG 8739

44 ysIleGlyAsp.HisHisThrHisGluHisAspGluSerValSerHisV 60
:: ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8740 AAGCGGACGATTAACCATACGCATACGCATACAGACTCCGCCCCAGTT..... 8782

60 aIglyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisHisAlaThr 76
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8783 .....TGGATACGACCAAGCGGAGCAGC 8806

77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln..... 88
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8807 AAGCGCAACAAGTACCGGTACATGCTTAAGCAGGATCACACCGTTA 8856

88 ..... 88

8857 AAGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGTGTAGA 8906

89 .....SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8907 AGSGCTTAGTACAAAGGATACCTTTCCTCGCAAAATGCCTCCACGGGA 8956

98 .....MetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.... 111
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8957 CAGCGTAAAGGTTAGCATAGTAGTAGTACACTCAGCAACGTCATGTACAC 9006

112 .....GlyHisGluProTyrGlnLeu 118
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGAGGGGAAAAAATATGATCTA 9056

119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
9057 CTTCCCGTTTAC.....GGTAAAGAANTTCCTTGGCACA...GTGTACGA 9097

135 nAspLeuLys...SerArgGlyIleLeu.....ProAla.Asn 146
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

alignment_scores:
  Quality: 81.00      Length: 210
  Ratio: 0.853      Gaps: 14
  Percent Similarity: 45.238      Percent Identity: 25.238

alignment_block:
US-09-674-779-2 x US-08-931-869-1 ..

Align seg 1/1 to: US-08-931-869-1 from: 1 to: 16656

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
||||| ||| ||| ::||| ||| |||
8649 ACCTTCACACGCCCTACTTGGCCACTCTGCTACTGCGCCACTATGA 8698

```



```

170 CCCAACCGCAAGATCGCCAAACCGCCACCCACCAACCGCCA..... 211
39 uileThrLysAspLysIleGlyAspHisThrHisGluHisAspGluS 56
212 .....CACCGCATCAGCGC. 226
56 erValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMet 72
227 .....TACCAGTTC 235
73 HisHisAlaThrLysGlnGluValValArgTrpGlnAlaTrpLeuGln 89
236 CATCACCCAGCAAAACACCAA...CCCGTCACCACTCATCACAGTG 282
89 rArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrAla 106
283 ACAG..... 286
106 rGSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProGluHis 122
287 .....CAAGTGCCCTCCCCACCAT 313
123 LeuTrpGlyGlnIleValProThrLeuHisLeuTyrGln...AspLeu 138
314 CTTCCACACCAAGCTTCCACCCAGCTTCCACTCCAGCAGCTAAGTCTCCA 363
138 sSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArg 155
364 TCGTCGTGAGCTGCTCCCT...CAGGCTCAAGCCCAAGCTCCCAACGGC 410
155 snProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHis 171
411 TGACGCTATTCTCCAAGTGGCACTCCGCCATCAGCGCGTTGCTATTG 460
172 ThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 188
461 .....CTGGAAGTG..... 469
188 nAlaLeuTyrGluLeuGlnAsnArgLeuGlyGlnTyrTrpLeuGluHis 205
470 .CTCTGTGCGAGTTTCTTCGCGATTGT.....TGCTTAGATT. 511
205 lyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGly 218
512 .....ATGGGATTTCCTTTTCGGGTTTCTCTATTGTT 544

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq;seq:US-08-203-532F-3
seq_documentation_block:
; Sequence 3, Application US/08203532F
; Patent No. 5856121
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; APPLICANT: Walsh, Kenneth
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2686
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,532F
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```



```

81  alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
337 CTGGCATAGCT.....CTGCTTCCA 358

98  MetSerGlnLeuThrThrAlaArgSerTrpGlnAla.....CysG1 112
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
359 .....GCCGACTCTGGAGGCGCCAGAGTTGG 387

112 yHisGluPro.....TyrGlnLeuProProGluH 122
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
388 GGACGACCCCGCGCTCTGTCTCCAACTCTTCCAGCTTGGGCTCCAGC 437

122 isLeu...TrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
438 ACCCGAGCTGGGCGCGTG..... 457

138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
458 CGCGCGGGGACTACGCGCGCCAGGCACTGTC..... 490

154 gAsnProGluLeuAsnGlnCysAlaGlyAlaAlaMetSerLysHisL 171
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
491 .....ACCTGGCGGCGGAGAGGAGCGAGCGCGCGGCAAG 524

171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
525 AGGAAAAGCGACAGCTCAGACTCCCGAGGAGAAATACAGTCAGAGT 574

188 Gln.....AlaLeuTyrGlnLeuGlnAsn.. 195
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
575 CAACAGCAAAACCCAGGAAAGAGGACAGCATTTACCAAGAGCAATCA 624

196 .....ArgLeuCysGlnTyrTrpLeuGluHis 204
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
625 GAGAACTTGAAGCAGAAATTTGCCCATCATTAATTCAC 664

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-01882A-3

seq_documentation_block:

```

; Sequence 3. Application PC/TUS9501882A
; GENERAL INFORMATION:
; APPLICANT: Gorski, David H.
; TITLE OF INVENTION: Growth Arrest Homeobox Gene
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Calfee, Halter, and Griswold
; STREET: 800 Superior Avenue
; CITY: Cleveland
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 44114-2688
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01882A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Goltick, Mary E.
; REGISTRATION NUMBER: 34829
; REFERENCE/DOCKET NUMBER: 22311/00114
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (216) 622-8200
; TELEFAX: (216) 241-0816
; TELEX: 980499
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 33..941
PCT-US95-01882A-3

```

alignment_scores:

```

Quality: 80.50 Length: 180
Ratio: 1.019 Gaps: 10
Percent Similarity: 43.889 Percent Identity: 21.667

```

alignment_block:

US-09-674-779-2 x PCT-US95-01882A-3 ..

Align seg 1/1 to: PCT-US95-01882A-3 from: 1 to: 941

```

48 HisHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAl 64
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
246 CACCACCACTCACCACCATCAGCAGCAGCAGCAGCAGCAGCAGCAGC 295

64 ahisPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValV 81
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
296 CAAC.....TGGCAGCTCCCGCAGATGCTTCCCCACCGAGTGCGG 336

81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
337 CTGGCATAGCT.....CTGCTTCCA 358

98 MetSerGlnLeuThrThrAlaArgSerTrpGlnAla.....CysG1 112
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
359 .....GCCGACTCTGGAGGCGCCAGAGTTGG 387

112 yHisGluPro.....TyrGlnLeuProProGluH 122
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
388 GGACGACCCCGCGCTCTGTCTCCAACTCTTCCAGCTTGGGCTCCAGC 437

122 isLeu...TrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu 137
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
438 ACCCGAGCTGGGCGCGTG..... 457

138 LysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrAr 154
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
458 CGCGCGGGGACTACGCGCGCCAGGCACTGTC..... 490

154 gAsnProGluLeuAsnGlnCysAlaGlyAlaAlaMetSerLysHisL 171
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
491 .....ACCTGGCGGCGGAGAGGAGCGAGCGCGCGGCAAG 524

171 euThrAsnSerAlaIleAspIleTrpValProAspLeuGluIleLysSer 187
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
525 AGGAAAAGCGACAGCTCAGACTCCCGAGGAGAAATACAGTCAGAGT 574

188 Gln.....AlaLeuTyrGlnLeuGlnAsn.. 195
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
575 CAACAGCAAAACCCAGGAAAGAGGACAGCATTTACCAAGAGCAATCA 624

196 .....ArgLeuCysGlnTyrTrpLeuGluHis 204
   :||||:||||:||||:||||:||||:||||:||||:||||:||||:
625 GAGAACTTGAAGCAGAAATTTGCCCATCATTAATTCAC 664

```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-524-2

seq_documentation_block:

```

; Sequence 2, Application US/09479524
; Patent No. 6268350
; GENERAL INFORMATION:
; APPLICANT: Barbera-Guillem, Emilio
; TITLE OF INVENTION: Polynucleotides For Inhibiting Metastasis And Tumor Cell Growth

```



```

602 .....CAGCTCTTCCACCTCGACGAGGAACTGGCAGAACTCCCGTGAGTTC 558
114 uProTyrGlnLeuProGluHisLeuTrpGlyGlnIleValProThr. 130
    |||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
557 ACCAACCAAGCCCTTAAAGTATCATCTTTTGAAGCAAAATACCAACCC 508
131 ..LeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla.As 146
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
507 CAGTACACCCCTCTGAAAAA...AAGAGCGAGGAGTGTGGCCCATTTAA 461
146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
460 C.....AGGAACCCCACTCAAGGTCATCCCT 432
163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
431 CTGGA.....ATGGAAGACACATATGGAACCGTCT...GATCTCTGG 391
179 ..... 179
390 AGTGAAGTATAAGAAAGGTGGCTTGTGATCAACGAAACTGGGTGTACT 341
180 .....ValProAspLeuGluIle..... 185
340 TCGTGATTCCAAAGTATACCTCCGGGGTCAGTCTTGCACAAACAGGCC 291
186 ..LysSerGlnAlaLeuTyrGlu..... 192
290 CTAACCAAGAGTCTATATGAGAACTCTAAGTATCTCTGAGATCTGGT 241
193 .....LeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisG1 205
240 GCTAATGGAGGAGAGAGAGGTGAAGTCTGCTACTGCGCCAGATATGG 191
205 yGluAsnGln..AsnPhgGlyLeuGlyLeuTyrAlaThrcllyAlaIleHis 221
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
190 CCCACAGCAGCTACCTGGGGGAGTATTCATCTTACCAGTCTGACCAT 141
222 Leu 222
140 TTA 138

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-846-762-1

```

seq_documentation_block:
; Sequence 1, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846, 762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 24417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-1

```

```

alignment_scores:
Quality: 80.00      Length: 364
Ratio: 0.526      Gaps: 19
Percent Similarity: 41.758      Percent Identity: 19.505

alignment_block:
US-09-674-779-2 x US-08-846-762-1/rev ..

```

```

Align seg 1/1 to reverse of: US-08-846-762-1 from: 1 to: 24417
1 MetLysAsnPhenAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
5860 TTGCGCTCGATCAACGAGCGGGGTGTAAACGTTGTGTAATACCATGCT 5811
17 uValAlaCysSerAlaPro..... 23
5810 CGGCCCGCAGAACACGCCCTCTTCGAGAGTGACATTGTTCATATACCGACA 5761
24 .....IleProThr 26
5760 CGTTGCTTCTGGATCTTCGAGCGATCACCATAACGACCTTGTTCGCCAGC 5711
27 .....AsnProGlnValSerProIleLysThrProSerVa 38
5710 AATAGTTTCTGACCCACGAAACCCCTGCGCAATCCGGGCACCTGCACA 5661
38 lleIleThrLys..AspLysIleGlyAspHisHisThrHisGluHisAsp 54
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
5660 GATGTGCACGAAAGTGCACAACTCGGGAGTCGTACCGATCTGCCACCGT 5611
55 GluSerValSerHisValGly.....LeuGlnAlaHi 65
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
5610 CGTCGACGATCGCGCTGGGTGCTGATAATAACTCATGCTGTGAGACTCA 5561
65 sPhcGluThrTrpLeu.....GlnM 72
5560 CACTCGATTGTGGTTGTTCAAACAGCGCTGCAATTGACGGCTCGCGAAA 5511
72 etHisHisAlaThrLysGlnCluValValArgTyrGlnAlaTyrLeuGln 88
    ::|||  ::  |||||  ::  |||||  ::  |||||  ::  |||||  ::
5510 TATCCATAGCGCGAGCCGCCACCTCGTTAAGTGATGGTATCTCTGGTA 5461
89 SerArgLeuGlyAsnTyrLeuProProMetSerGlnLeuLeuThrAl 105
5460 AATAAGCTC.....TCCGTCTGTATAGGAG 5435
105 aArgSerTrpGlnAlaCysGly.....HisGluProTyrGlnLeuP 119
5434 CATCGCGAAGAGGCGCTGGAAGAAAAATCCATGAACCGTAC..... 5392
119 roProGluHisLeuTrpGly.....GlnIleVal 128
5391 .....GTTTGAATACCTCGGCTAGAGCAGCAGCAACAGCTGGTT 5353
129 ProThrLeuHisLeuTyrGlnAspLeuLysSer...ArgGlyIleLeuPr 144
5352 CGCACCTTGGCAGCTGGAGCTAACAAGCGAGCTGACAAACCCAGCTCCC 5303
144 oAlaAsnThrGlnIleArgSerVal.....TyrArgAsnP 156
5302 TGAACGACGACACAGCTGCACCTTTGATTTCGAACATCGGTATCTGCCCG 5253
156 roGluLeuAsnGlnCysAlaGlyAlaAlaAlaMetSerLysHisLeuThr 172
5252 AGTAGACGACCTGCTTGCCTGCGCATAGCTGATCGCAAGGAGGCACG 5203
173 AsnSer.....AlaIleAspIleTrpValPro..... 181
5202 CATAGCTGGGCAAAATCGCGACTCTGCATCTGTACTGCCACATGCCCG 5153
182 .....AspLeuGluIleLysSerGln..... 188
5152 CCACAATGATCTT...GTCAAGTTTCAACATTTGATTGGGCAACTGCATCA 5106
189 .....AlaLeuTyrGluLeu 193
5105 ATCTGCGCTTGGCAGGTTTCCCGCCCCAACGGGCAACGTTCAAGATC 5056
194 GlnAsnArgLeuCysGlnTyrTrp..... 201

```

5055 GAAACGAG.....TATTGGCACACACTGCTTCCGGTGAGTACTC 5015
 202 LeuGluHis..... 204
 : : : : :
 5014 GGTAAACACACACTGACTCGTGTGCAACACCGCTCGAAGAAGTAGTTAAGC 4965
 205GlyGluAsnGlnAsnPheGlyLeuGlyLeuTyr 215
 : : : : :
 4964 TCCGCAGCATGGCTATCACCAGTACCAGACACACGGGGTACGGCATTCAC 4915
 216 AlaThrGlyAlaIle.....HisLeuAspThrGlnGlyPheArgLysTr 230
 : : : : :
 4914 GCTGCTCGCTTGCACTCCCTACCTGACCAACACCATGGCAGATCAGTCTG 4865
 230 pGlyAlaGlnPheSerGluThrAsnSerIleCysArgHis 243
 : : : : :
 4864 GTACAGCGTAGCGCTCAACTGAATGGCGCGCGGCCAC 4825

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-440-845D-8

seq_documentation_block:

; Sequence 8, Application US/08440845D
 ; Patent No. 595329
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuan, L.
 ; APPLICANT: Kridl, J.
 ; APPLICANT: Dehesh, K.
 ; APPLICANT: Knauf, V.
 ; TITLE OF INVENTION: Engineering Plant Thioesterases For
 ; TITLE OF INVENTION: Altered Substrate Specificity.
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Calgene, Inc.
 ; STREET: 1920 Fifth Street
 ; CITY: Davis
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 95616
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 7.0
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/440,845D
 ; FILING DATE: 15-MAY-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Donna E. Scherer
 ; REGISTRATION NUMBER: 34,719
 ; NAME: Carl J. Schwedler
 ; REGISTRATION NUMBER: 36,924
 ; REFERENCE/DOCKET NUMBER: CGNE 113
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (916) 753-6313
 ; TELEFAX: (916) 753-1510
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1300 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; US-08-440-845D-8

alignment_scores:
 Quality: 79.50 Length: 244
 Ratio: 0.787 Gaps: 13
 Percent Similarity: 41.393 Percent Identity: 22.951
 alignment_block:
 US-09-674-779-2 x US-08-440-845D-8/rev ..

Align seg 1/1 to reverse of: US-08-440-845D-8 from: 1 to: 1300

32 ProIleLysThrProSerValLeuIleThrLysAspLysIleGlyAspHis 48
 : : : : :
 645 CCTCTGGAATCCCAATCGAGTGTCTTTGGACAGTGAACCAAGTACTCA 596
 : : : : :
 48 sHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHis 65
 : : : : :
 595 TCACGAACATCAACAT.....CGAC 576
 65 isPheGlu...ThrTrpLeu..... 70
 : : : : :
 575 TTTTGAAGTCGCTGCTGCTTGTTCATCATTACCCACTTGCCTAGTAG 526
 : : : : :
 71GlnMetHisHisAlaThrLysGlnGluValVa 81
 : : : : :
 525 CTCGGCCATAACTTCACAGTGGGATAGTCTCTCAGAAATCCATCATCGT 476
 81 lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPom 98
 : : : : :
 475 CTGGTTCGGATTTTCTTCACCTGGCCCGACGACTCTATTTCACCCAC 426
 : : : : :
 98 etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys GlyHisGln 114
 : : : : :
 425 ATCACTCCCAAGCTGGATATTGTAGATTTCGATGTCATCGCAGCAGTAA 376
 : : : : :
 114 uProTyrGlnLeuProProGluHisLeuTrpGlyGlnIleValProThr. 130
 : : : : :
 375 CCCATATCAGACGCAATTTTCTCATGTAGGGGTGTGCGAAACCCACCC 326
 : : : : :
 131LeuHisLeuTyrGlnAsp 136
 : : : : :
 325 GTCGAATATCAACAGCTTTGGGGGTGATTCATCCAACTCTCTGCAAGAG 276
 : : : : :
 137LeuLysSerArgGlyIleLeuProAlaAsnThrGlnI 149
 : : : : :
 275 GTTGGCAATAGTCTCAACAGTAGCGGTCTTTGTAATCCCAACCTC...AT 229
 : : : : :
 149 eArgSerValTyr.....ArgAsnProGluLeuLeuAsnGlnC 161
 : : : : :
 228 AGCATCTCACTATGAACTTCTCTTATACGAAAGCCCTCTCGGTCAAG 179
 : : : : :
 161 yAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp 177
 : : : : :
 178 CTGCCAGCGGAGCGCGTCAGCCAGCCCGGAGAGAAC..... 141
 : : : : :
 178 IleTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuG 194
 : : : : :
 141 141
 194 nAsnArgLeuCysGlnTyrTrpLeuGluHis.....GlyGluAsnG 208
 : : : : :
 140 .CACCGCTCTGTTTGTGATGATTAACTTGTCTGATGAGGAGGAGAAACC 92
 : : : : :
 208 lAsnPhe.....Gly LeuGlyLeuTyrAlaThrGlyAlaI 220
 : : : : :
 91 ACAATTATCCGGGAGGAATGGAGTTGGGCTG.....GCCGGGT 51
 : : : : :
 220 eHisLeuAspThrGlnGlyPheArgLys 229
 : : : : :
 50 GGAATGCGGGCCAATGGCTTCGGGAA 23

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-868-458-8

seq_documentation_block:

; Sequence 8, Application US/08868458
 ; Patent No. 6150512
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuan, L.
 ; TITLE OF INVENTION: Engineering Plant Thioesterases And
 ; Disclosure of Plant Thioesterases
 ; Having No. 6150512el Substrate Specificity

;
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Calgene, Inc.
 ; STREET: 1920 Fifth Street
 ; CITY: Davis
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 95616
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
 ; COMPUTER: IBM PC
 ; OPERATING SYSTEM: Windows NT 4.0
 ; SOFTWARE: Microsoft Word For Window 95 7.0a
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/868,458
 ; FILING DATE: 03-Jun-1997
 ; CLASSIFICATION: <Unknown>
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/07064
 ; FILING DATE: 15-MAY-96
 ; APPLICATION NUMBER: 08/537,083
 ; FILING DATE: 29-SEPT-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Carl J. Schwedler
 ; REGISTRATION NUMBER: 36,924
 ; REFERENCE/DOCKET NUMBER: CGNE 133
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (916) 753-6313
 ; TELEFAX: (916) 753-1510
 ;
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1300 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ;
 ; MOLECULE TYPE: cDNA to mRNA
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 ; US-08-868-458-8

alignment_scores:
 Quality: 79.50 Length: 244
 Ratio: 0.787 Gaps: 13
 Percent Similarity: 41.393 Percent Identity: 22.951

alignment_block:

US-09-674-779-2 x US-08-868-458-8/rev ..

Align seg 1/1 to reverse of: US-08-868-458-8 from: 1 to: 1300

```

32  ProLeuThrProSerValLeuLeuThrLysAspLysIleGlyAspHi 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
645  CCTGTGAAATGCCAATCTGAGTTCTCTGGACAGTGAACCAAGTACTCA 596

48  sHisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAla 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
595  TCAGAACATCAACAT.....CGAC 576

65  isPheGlu...ThrTrpLeu..... 70
|||||:|||||:|||||:|||||:|||||:|||||:
575  TTTTGAAGTCGCTGGTGTCTTCATCATCCACTTCTAGTAG 526

71  .....GlnMetHisHisAlaThrLysGlnGluValVa 81
|||||:|||||:|||||:|||||:|||||:|||||:
525  CTCGGCAATAACTTGACCAAGTGGCAGTCTCTCAGAATCCAATCACGT 476

81  laArgThrGlnAlaThrLeuGlnSerArgLeuGlyAsnTrpLeuProPom 98
|||||:|||||:|||||:|||||:|||||:|||||:
475  CTGTTCCGATTTTCTTCACCTGGCCCAAGACTCTATTTCACAC 426

98  etSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCys.GlyHisG 114
|||||:|||||:|||||:|||||:|||||:|||||:
425  ATCACTCAAGCTGGATATTGTGATTTTCGATGTCATCGGAGCAGTAA 376

```

```

114  uProTyGlnLeuProGluHisLeuTrpGlyGlnIleValProThr. 130
|||||:|||||:|||||:|||||:|||||:|||||:
375  CCCATATCAGACGCAATTTTCTCATGGTGGGTGTGCGAAACCCACCC 326

131  .....LeuHisLeuTyGlnAsp 136
|||||:|||||:|||||:|||||:|||||:
325  GTCGAATATCCAAAGCTTTGGGGGTGATTGCATCCCAACCTCTGCAAGAG 276

137  .....LeuLysSerArgGlyIleLeuProAlaAsnThrGlnI 149
|||||:|||||:|||||:|||||:|||||:
275  GTTGCAATAGTCTCAACAGTAGCGGTCTGTGTAATCCCAACTC...AT 229

149  eArgSerValTy.....ArgAsnProGluLeuAsnGlnC 161
|||||:|||||:|||||:|||||:|||||:
228  AGCATCTCATATGAACCTTCTCTTATACGAAAGCCGCTCGGTCAAG 179

161  ysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp 177
|||||:|||||:|||||:|||||:|||||:
178  CTGCCAGCGGAGCGGTACGCCAGCCGCCGAGAAAC..... 141

178  ileTrpValProAspLeuGluIleLysSerGlnAlaLeuTyGluLeuG 194

141  ..... 141

194  nasnArgLeuCysGlnTyTrpLeuGluHis.....GlyGluAsnG 208
|||||:|||||:|||||:|||||:|||||:
140  .CACGCCCTCTGTTTGTAGTGAGTAAACCTTGCTGATGAGGAGAAACC 92

208  InAsnPhe.....GlyLeuGlyLeuTyAlaThrGlyAlaI 220
|||||:|||||:|||||:|||||:|||||:
91  ACAATTATCCGGGAGGAGTGGCTGGCCCTG.....GGCCGGGT 51

220  eHisLeuAspThrGlnGlyPheArgLys 229
|||||:|||||:|||||:|||||:
50  GGAATGCGGCGCAATGGGCTTCGGGAA 23

```

seq_name: /cgn_2_6/ptodata/2/ina/5A_COMB.seq:US-08-801-263A-8

seq_documentation_block:

; Sequence 8, Application US/08801263A
 ; Patent No. 5811407
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnston, Robert E.
 ; APPLICANT: Davis, Nancy L.
 ; APPLICANT: Simpson, Dennis A.
 ; TITLE OF INVENTION: System for the In Vivo Delivery and
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
 ; STREET: 1211 East Morehead Street
 ; CITY: Charlotte
 ; STATE: No. 5811407th Carolina
 ; COUNTRY: USA
 ; ZIP: 28234
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/801,263A
 ; FILING DATE: 19-FEB-1997
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sibley, Kenneth D.
 ; REGISTRATION NUMBER: 31,665
 ; REFERENCE/DOCKET NUMBER: 5470-147
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-420-2200
 ; TELEFAX: 919-861-3175
 ; INFORMATION FOR SEQ ID NO: 8:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-801-263A-8

alignment_scores:
  Quality: 79.50      Length: 211
  Ratio: 0.837       Gaps: 14
  Percent Similarity: 45.024  Percent Identity: 25.118

alignment_block:
US-09-674-779-2 x US-08-801-263A-8 ..

  Align seg 1/1 to: US-08-801-263A-8 from: 1 to: 11703

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
||||| ||| ||| :||| ||| |||
8649 ACCCTGACGAGCCCTACTTGGGCACATGCTGCTACTGCCACCATCTGA 8698

27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
||| ||| ||| ||| ||| ||| ||| |||
8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.....GAGC 8739

44 yslleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisv 60
:: ||| ||| ||| ||| ||| ||| ||| |||
8740 AAGCGGACGATAACACCATACGATACAGACTTCCGCCCCAGTT..... 8782

60 alGlyLeuGlnAlaHisPheGluThrTripleuGlnMethHisAlaThr 76
||| ||| ||| ||| ||| ||| ||| |||
8783 .....TGGATACGACCAAGCGGAGCAGC 8806

77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln..... 88
||| ||| ||| ||| ||| ||| ||| |||
8807 AAGCGCAACAAAGTACGCTACATGTCGTTGACGAGGATCACCGTTA 8856

88 ..... 88

8857 AAGAGGACCATGGATGACATCAAGATTAGCACCCTCAGGACCGGTGTA 8906

89 .....SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
||| ||| ||| ||| ||| ||| ||| |||
8907 AGGCTTAGCTACAAAGGATATCTTCTCTCGCAAAATGCCCTCCAGGGGA 8956

98 ....MetSerGlnLeuLeuThrThrAlaAlaArgSerTrpGlnAlaCys.... 111
::: ||| ||| ||| ||| ||| ||| ||| |||
8957 CAGCGTAACGGTTAGCATAGTAGTAGTCAACTCAGCAACGTCATGTACAC 9006

112 .....GlyHisGluProTyrGlnLeu 118
||| ||| ||| ||| ||| ||| ||| |||
9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAATATGATCTA 9056

119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
||||| ||| ||| ||| ||| ||| ||| |||
9057 CCTCCCGTTCAC.....GGTAAAAAATTCCTTGACACA...GTGTACGA 9097

135 nAspLeuLys.....SerArgGlyIleLeu.....ProAla. 145
: ||| ||| ||| ||| ||| ||| ||| |||
9098 CCGTCTGAAGAAACAACTGCGAGGTACATCATCTATGCACAGGCCCGGAC 9147

146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
::: ||| ||| ||| ||| ||| ||| ||| |||
9148 CGCAGCTTATACATCTACCTACCTGGAAGATCAACAG..... 9183

162 aGlyGlyAlaAlaMetSerLysHisLeu 171
||| ||| ||| ||| ||| ||| ||| |||
9184 .GGAAGTTTACGAAAGCCGCCATCTG 9210

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-09-102-248-8
```

```
seq_documentation_block:
; Sequence 8, Application US/09102248
; Patent No. 6008035
; GENERAL INFORMATION:
; APPLICANT: Johnston, Robert E.
; APPLICANT: Davis, Nancy L.
; APPLICANT: Simpson, Dennis A.
; TITLE OF INVENTION: System for the In Vivo Delivery and
; TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 6008035th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,248
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/801,263
; APPLICATION NUMBER: 19-FEB-1997
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5470-147
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-102-248-8

alignment_scores:
  Quality: 79.50      Length: 211
  Ratio: 0.837       Gaps: 14
  Percent Similarity: 45.024  Percent Identity: 25.118

alignment_block:
US-09-674-779-2 x US-09-102-248-8 ..

  Align seg 1/1 to: US-09-102-248-8 from: 1 to: 11703

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
||||| ||| ||| :||| ||| |||
8649 ACCCTGACGAGCCCTACTTGGGCACATGCTGCTACTGCCACCATCTGA 8698

27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
||| ||| ||| ||| ||| ||| ||| |||
8699 ACCGTGCTTCAGCCCTGTTAAGATCGAGCAGGTCTGG.....GAGC 8739

44 yslleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisv 60
:: ||| ||| ||| ||| ||| ||| ||| |||
8740 AAGCGGACGATAACACCATACGATACAGACTTCCGCCCCAGTT..... 8782

60 alGlyLeuGlnAlaHisPheGluThrTripleuGlnMethHisAlaThr 76
||| ||| ||| ||| ||| ||| ||| |||
8783 .....TGGATACGACCAAGCGGAGCAGC 8806

77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln..... 88
||| ||| ||| ||| ||| ||| ||| |||
8807 AAGCGCAACAAAGTACGCTACATGTCGTTGACGAGGATCACCGTTA 8856

88 ..... 88

8857 AAGAGGACCATGGATGACATCAAGATTAGCACCCTCAGGACCGGTGTA 8906

89 .....SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
||| ||| ||| ||| ||| ||| ||| |||
8907 AGGCTTAGCTACAAAGGATATCTTCTCTCGCAAAATGCCCTCCAGGGGA 8956

98 ....MetSerGlnLeuLeuThrThrAlaAlaArgSerTrpGlnAlaCys.... 111
::: ||| ||| ||| ||| ||| ||| ||| |||
8957 CAGCGTAACGGTTAGCATAGTAGTAGTCAACTCAGCAACGTCATGTACAC 9006

112 .....GlyHisGluProTyrGlnLeu 118
||| ||| ||| ||| ||| ||| ||| |||
9007 TGGCCCGCAAGATAAAACCAAAATTCGTGGGACGGGAAAATATGATCTA 9056

119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGl 135
||||| ||| ||| ||| ||| ||| ||| |||
9057 CCTCCCGTTCAC.....GGTAAAAAATTCCTTGACACA...GTGTACGA 9097

135 nAspLeuLys.....SerArgGlyIleLeu.....ProAla. 145
: ||| ||| ||| ||| ||| ||| ||| |||
9098 CCGTCTGAAGAAACAACTGCGAGGTACATCATCTATGCACAGGCCCGGAC 9147

146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
::: ||| ||| ||| ||| ||| ||| ||| |||
9148 CGCAGCTTATACATCTACCTACCTGGAAGATCAACAG..... 9183

162 aGlyGlyAlaAlaMetSerLysHisLeu 171
||| ||| ||| ||| ||| ||| ||| |||
9184 .GGAAGTTTACGAAAGCCGCCATCTG 9210
```

```

|||||..... ||| |||||..... |||
8807 AACGCAACAAGTACCGTATACGTGCGCTTGAGCAGGATCACACCGTTA 8856
88 .....
8857 AGAAGGCACCATGGATGACATCAAGATTAGCACCTCAGGACCGGTGAGA 8906
89 .....SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
8907 AGCCTTAGCTACAAGAGTACTTCTCTCCGCAAAATGCCCTCCAGGGGA 8956
98 ....MetSerGlnLeuLeuThrAlaArgSerTyrGlnAlaCys..... 111
8957 CAGCGTAACGGTTAGCATAGTAGTAGCAACTCACCAACGTCATGTACAC 9006
112 .....GlyHisGluProTyrGlnLeu 118
9007 TGGCCCGCAAGATAAACCAAAATTCGTGGGACGGGAAAAATATGATCTA 9056
119 ProProGluHisLeuTyrGlnIleValProThrLeuHisLeuTyrGln 135
9057 CCTCCCGTTCAC.....GGTAAAAAAATTCCTTCACACA...GTACGA 9097
135 nAspLeuLys.....SerArgGlyIleLeu.....ProAla. 145
9098 CGGTCGTAAGAAACAACATGCGAGGCTACATCACTATGCACAGCGCGGAC 9147
146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
9148 CGCAGCTTATACATCTCTACCTCGGAAGATCATCAG..... 9183
162 aGlyGlyAlaLaMetSerLysHisLeu 171
9184 .GGAAGTTTACGCAAGCGGCATCTG 9210

seq_name: /cgn2_6/ptodata/2/ina/PCtUS_COMB.seq:PCT-US95-00362-4

seq_documentation_block:
; Sequence 4, Application PC/TUS9500362
; GENERAL INFORMATION:
; APPLICANT: IMMUNEX CORPORATION
; TITLE OF INVENTION: Ligand That Binds Fas Antigen
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: US
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/00362
; FILING DATE: 06-JAN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,138
; FILING DATE: 07-JAN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/190,559
; FILING DATE: 01-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Anderson, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2805-WO
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 base pairs
; TYPE: nucleic acid

```

```

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..870
PCT-US95-00362-4

alignment_scores:
  Quality: 79.00 Length: 250
  Ratio: 0.814 Gaps: 15
  Percent similarity: 38.800 Percent Identity: 25.200

alignment_block:
US-09-674-779-2 x PCT-US95-00362-4 ..
Align seg 1/1 to: PCT-US95-00362-4 from: 1 to: 1171
23 ProfileProThrAsnProGlnValSerProIleLysThrProSer..... 37
163 CGCGCACTCCACACCACTGTCACCACTACCAACCGCATCACACCA 212
38 .....ValLeuIleThrLysAspLysIleGlyAspHisHisThrH 51
213 ACTCCCCACTGTCGCCACTGACCCCTCTAAAGAAAGAGCACACACAA 262
51 isGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGlu 67
263 AT..... 264
68 ThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
265 CTGTGGCTACCG.....GTGCTATTTT 287
84 nAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProMetSerGlnL 101
288 CATGGTCTCGTGGTGTGTTGGAATGGGATTA...GGAATGTATCAGC 334
101 euLeu.ThrThrAlaArgSerTyrGlnAlaCysGlyHisGluProTyrG 117
335 TCTTCACCTGCAGAAAGGAACTGGCAGAACTCGTGAGTTCACCAACCA 384
117 nLeuProProGluHisLeuTyrGlyGlnIleValProThr....LeuHisL 133
385 AGCCTTAAAGTATCATCTTTTGAAGACAAATAGCCACCCAGTACACC 434
133 euTyrGlnAspLeuLysSerArgGlyIleLeuProAla.AsnThrGlnI 149
435 CTCTGAAAAA...AAGACCGAGGAGTGTGGCCCATTTAAC..... 473
149 eArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAla 166
474 .....AGGAAACCCCACTCAAGGTCCCATCCCTCTGGA.... 506
166 laMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp..... 179
507 ..ATGGGAAGACACATATGAACCGCTCT...GATCTCTGGAGTGAAGTA 551
179 ..... 179
552 TAAGAAAGGTGGCCTTGTCATCAACGAAACTGGGTGTGACTTCGTGTATT 601
180 .....ValProAspLeuGluIle.....LysSerG 188
602 CCAAAGTATACTTCGGGGTTCAGTCTTGCAACACACGAGCCCTTAACAC 651
188 laAlaLeuTyrGlu..... 192
652 AAGGCTCTATATGAGAACTCTAAGTATCCTGAGGATCTGGTGCTAATGGA 701

```

alignment_scores:		
Quality:	78.00	Length:
Ratio:	0.857	Gaps:
Percent Similarity:	45.050	Percent Identity:
		24.257

alignment_block:

US-09-674-779-2 x US-08-458-120-1 ..

Align seg 1/1 to: US-08-458-120-1 from: 1 to: 1882

```

82 ArgTyrGln.....AlaTyrLeuGln.....SerAr 90
889 AGATATCAGCCAGCATGTGAAGTTTATCCAGGTTGATATCTGTGCAGAGA 938
90 gLeuGlyAsnTyrLeuProPomMetSerGlnLeu..... 102
939 ATTGGGAATAAGTAAGACCGCGCTTACTTGTAGAAACATACATG 988
103 ..ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu 118
989 CTGTCATAAGCAGCTTTAGAGGAACCTGATAAACACCATCGGAGTAT 1038
119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHis..... 132
1039 CCTCCAGAGAGCAAGTGGTGGAAA.....ACTCTGAGAGAAAAAT 1079
133 .....LeuTyrGlnAspLeuLysSerArgGlyIleLeu 144
1080 GAGAGCAATGAAGTCATCAAGGAAGTACTTCTAAAAATCCCTGC 1129
144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu...Asn 159
1130 CTATGAATTATTACACAGATTCTTACCATGTTCAAGAACACACTACCTAGA 1179
160 GlnCys.....AlaGlyGlyAlaAlaMetSerLysHisLeuThrAs 173
1180 GACTGTTTCGTGGTAAGTGAAGGAGCA.....GGAC 1225
1207 .AATACTATGGACATT.....GGAC 1225
190 euTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGluHisGlyGlu 206
1226 GGACTGTGCTTCAAGAACTACCTTCTCGTCACAGGCTTGATCTGTGACT 1275
207 AsnGlnAsnPheGlyLeuGlyLeu.....TyrAlaThrGlyAlaI 220
1276 TTCGGAACAATGGAGTGGTGGATTGCTATTGTCAGCTGCGGTGT 1325
220 eHisLeuAsp.....ThrGlnGlyPheA 228
1326 GCCTAAGATAGAGCCCTGGCAATGGATCATCTGTGGGAAGGAGACA 1375
228 rgLysTrpGlyAlaGlnPheSerGluThrAsnSerIleCysArgHisVal 244
1376 GTGCATTGGGTTTCTGCGATGGAGGTAGAAACCATCTGCAGGTACAC 1425
245 LeuPro 246
1426 TTGCCA 1431

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-867-970-1

seq_documentation_block:

Sequence 1, Application US/08867970

Patent No. 5945273

GENERAL INFORMATION:

APPLICANT: Olsen, Henrik S.

APPLICANT: Coleman, Timothy

TITLE OF INVENTION: Human Oxalyl-CoA Decarboxylase

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

```

COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,970
FILING DATE: JUN-3-97
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/458,120
FILING DATE: 02-JUN-1995
PRIOR APPLICATION DATA: WO PCT/US94/05561
APPLICATION NUMBER: 18-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Robert H.
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PF119DI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1882 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 10..1744
US-08-867-970-1

```

alignment_scores:

Quality: 78.00 Length: 202
Ratio: 0.857 Gaps: 11
Percent Similarity: 45.050 Percent Identity: 24.257

alignment_block:

US-09-674-779-2 x US-08-867-970-1 ..

Align seg 1/1 to: US-08-867-970-1 from: 1 to: 1882

```

82 ArgTyrGln.....AlaTyrLeuGln.....SerAr 90
889 AGATATCAGCCAGCATGTGAAGTTTATCCAGGTTGATATCTGTGCAGAGA 938
90 gLeuGlyAsnTyrLeuProPomMetSerGlnLeu..... 102
939 ATTGGGAATAAGTAAGACCGCGCTTACTTGTAGAAACATACATG 988
103 ..ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu 118
989 CTGTCATAAGCAGCTTTAGAGGAACCTGATAAACACCATCGGAGTAT 1038
119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHis..... 132
1039 CCTCCAGAGAGCAAGTGGTGGAAA.....ACTCTGAGAGAAAAAT 1079
133 .....LeuTyrGlnAspLeuLysSerArgGlyIleLeu 144
1080 GAGAGCAATGAAGTCATCAAGGAAGTACTTCTAAAAATCCCTGC 1129
144 roAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeu...Asn 159
1130 CTATGAATTATTACACAGATTCTTACCATGTTCAAGAACACACTACCTAGA 1179
160 GlnCys.....AlaGlyGlyAlaAlaMetSerLysHisLeuThrAs 173

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 10..1744
US-09-326-217-1

alignment_scores:
    Quality: 78.00      Length: 202
    Ratio: 0.857        Gaps: 11
    Percent Similarity: 45.050      Percent Identity: 24.257

alignment_block:
US-09-674-779-2 x US-09-326-217-1 ..

Align seg 1/1 to: US-09-326-217-1 from: 1 to: 1882

      82 ArgTyrGln.....AlaTyrLeuGln.....SerAr 90
      |||||
      889 AGATATCAGCCAGATGTGAAGTTTATCCAGTTGATATCTGTGCAAGA 938
      :|||
      90 gLeuGlyAsnTyrLeuProPomSelSerGlnLeu..... 102
      :|||
      939 APTGGGGAATATGAAAGCCGCGCTTTACTTTGCTTAGGAACATACATG 988

```

103 ..ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeu 118

```

989 CTGTCACCTAAGCAGCGTTTTAGAGGAACCTTGATAAAACACCATGGCAGTAT 1030
119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHis..... 132
||||| ||| :: ||||::
1039 CCTCCAGACGCAAGTGGTGGAAA.....ACTCTGAGAGAAAAAT 1079
133 .....LeuTy rGlnAspLeuLysSerArgGlyIleLeuP 144
::: ||| ||||:: ||||
1080 GAAGAGCATGAAGCTGCATCCAAGAAGACTAGCTTCTAAANAATCCCTGC 1129
144 roALaasnThrGlnIleArgSerValTy rGAsnProGluLeu...Asn 159
|| ||| : : : : : ||| : : :
1130 CTTATGAATTATTACACAGPATTTCTACCATGTTC AAGAACAAC TACCTAGA 1179
160 GlnCys.....AlacGlyAlaAlaMetSerLysHisLeuThraS 173
::: ||| |||||
1180 GACTGTTTCGTGGTAAGTAGGAGGACA..... 1206
173 nSerAlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaL 190
: : : : : ||||| : : :
1207 .AATACTATGGACATT.....GGAC 1225
190 euTy rGluLeuGlnAsnArgLeuCysGlnTy rTrpLeuGluHisGlyGlu 206
||||| ||| : : : : : |||||
1226 GGACTGTGTTCTCAGAACA CTACCTTCCTCGTCAGAGGCTTGATGCTGTACT 1275
207 AsnGlnAsnPheGlyLeuGlyLeu.....Ty rAlaThrGlyAlaIl 220
: : : : : ||||| : : :
1276 TTCGGAACAATGGGAGTGGTTGGATTGGATTGCTATTGCAGCTGCCGTGGT 1325
220 ehISLeuAsp.....ThrGlnGlyPhea 238
: ||| : : : : :
1326 GGCTAAAGACTAAGCCCTGGCAATGGATCATCTGTGTGGAAGGAGACA 1375
228 rdLy rTrpGlyAlaGlnPheSerGluThraSnSerIleCysArgHisVal 244
: : ||| : : : : : ||||| : : :
1376 GTGCATTGGGTTTCTGGCATGGAGGTAGAAACCATCTGCAGGTACAAC 1425
245 LeuPro 246
|||||
1426 TTGCCA 1431
seq_name: /cgm2_6/ptodata/2/1na/6B_COMB.seq:US-08-972-218-1
seq documentation block:

```


; Sequence 1, Application US/08972218
; Patent No. 6197502
; GENERAL INFORMATION:
; APPLICANT: Renner, Wolfgang A.
; APPLICANT: Orberger, Georg H.
; APPLICANT: Koller, Daniel
; APPLICANT: Bailey, James E.
; TITLE OF INVENTION: EXPRESSION CLONING PROCESSES FOR THE
; TITLE OF INVENTION: DISCOVERY, CHARACTERIZATION AND ISOLATION OF GENES
; TITLE OF INVENTION: ENCODING POLYPEPTIDES WITH A PREDETERMINED PROPERTY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,218
; FILING DATE: 14-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 008358-0004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13905 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-972-218-1

alignment_scores:
Quality: 77.50 Length: 211
Ratio: 0.816 Gaps: 14
Percent Similarity: 45.024 Percent Identity: 25.118

alignment_block:

US-09-674-779-2 x US-08-972-218-1 ..

Align seg 1/1 to: US-08-972-218-1 from: 1 to: 13905

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
||||| ||| ||| :||| ||| ||| :|||
8916 ACCCTGACCAACCCCTACTGTGGCAGCATGCTACTGCCACCATACGA 8965
27 nProGlnValSerProIleLysThrProSerValLeuIleThrLysAspL 44
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
8966 ACCGTGCTTCAGCCCTGTAAAGATCCAGCAGGTCTGG.....GACG 9006
44 ystleGlyAsp..HisHisThrHisGluHisAspGluSerValSerHisV 60
: : ||| ||| ||| ||| ||| :||| :||| :||| :||| :|||
9007 AAGCGGACGATAACACCATACGATACGACTTCCGCCCGAGTT..... 9049
60 alGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaThr 76
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9050TGGATACGACCATAGCGGAGCAGC 9073

77 LysGlnGluValVal.ArgTyrGlnAlaTyrLeuGln..... 88
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9074 AAGCGCAACAAGTACCGCTACATGTGCTTAAGCAGGATCACACCGTTA 9123
88 88
9124 AAGAAGGCACCATGGATGACATCAAGATTAGCACCCTCAGGACCGTGTAGA 9173
89SerArgLeuGlyAsnTyrLeu.....ProPro..... 97
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9174 AGGCTTAGCTACAAAGGATCTTCTCCTCGCAAAATGCGCTCCAGGGGA 9223
98MetSerGlnLeuLeuThrAlaArgSerTrpGlnAlaCys.... 111
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9224 CAGCGTAACGGTTAGCATAGTAGTAGCAACTCAGCAACGTCATGTACAC 9273
112GlyHisGluProTyrGlnLeu 118
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9274 TGGCCCGCAAGATAAAACCAAAATTCGTGGGAGCGGGAATAATATGATCTA 9323
119 ProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTrpG 135
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9324 CCTCCCGTTCAC.....GGTAAAAAAATTCCTTGACACA...GTGTACGA 9364
135 nAspLeuLys.....SerArgGlyIleLeu.....ProAla. 145
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9365 CCGTCTGAAGACAACAACACTGCAGGCTACATCACTATGCACAGCGCGGAC 9414
146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9415 GGCACGCTTATACATCCTCTACTGGAAGAATCATCAG..... 9450
162 alGlyAlaAlaMetSerLysHisLeu 171
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9451 .GGAAGTTTACGCAAGCGCGCATCTG 9477

OM of: US-09-674-779-2 to: EST:* out_format : pfs

Date: Sep 16, 2002 6:20 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlh
-Q-/cgn2_1/USFIO_spool/US09674779/runat_12092002_124121_25821/app_query.fasta_1.310
-DB-EST -QFMT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-DELEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -STAR=1 -MATRIX=blosum62 -TRANS=human40.coi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09674779_@CGLN_1_2604
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-674-779-2

Query length: 250

Database: EST*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 1624.540000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_gss:AG078683	-	110.50	195.32	0.4476	645	AG078683 Pan troglodytes DNA, C
gb_gss:AG078683	+	105.50	176.42	1.40	640	AG078683 IM0498P02F Mouse 10kb
gb_gss:AG078683	-	103.50	172.97	2.18	632	AG078683 IM0498P02F Mouse 10kb
gb_gss:AG078683	+	98.50	161.16	9.92	817	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	98.50	156.28	18.54	1270	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	98.50	152.90	7.93	643	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	97.00	151.49	9.51	621	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	97.00	137.72	12.75	764	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	96.00	136.72	17.53	874	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	95.00	162.26	8.62	454	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	95.00	155.47	20.59	774	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	95.00	154.99	21.88	808	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	94.50	157.40	16.07	599	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	94.50	157.29	16.29	605	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	93.50	153.88	25.22	700	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	92.50	140.83	134.49	1939	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	92.00	152.34	30.76	631	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	91.50	155.42	20.71	440	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	91.50	151.45	34.44	630	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	91.00	147.02	60.80	941	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	91.00	149.83	42.43	673	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	91.00	146.79	62.64	886	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	91.00	146.14	68.12	940	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.50	149.30	45.42	651	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.50	149.21	45.92	656	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.50	147.47	57.41	768	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.50	146.92	61.58	807	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.50	146.82	62.45	815	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.50	143.96	90.01	1055	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.00	149.41	44.76	594	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.00	147.60	56.49	700	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	90.00	146.89	61.82	746	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	89.00	148.81	48.34	533	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	89.00	148.46	50.54	550	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	89.00	146.67	63.61	647	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	89.00	141.61	121.73	1023	AG078683 SP_0132 AL H03 T7A St
gb_gss:AG078683	+	89.00	141.51	123.25	1032	AG078683 SP_0132 AL H03 T7A St

gb_est2:BI993724 + 88.50 148.96 47.45 485 ! BI993724 1031003D12.x2 C. re
gb_est2:BI993724 + 88.50 148.42 50.81 509 ! BI993724 1031003D12.x2 C. re
gb_est2:BI993724 + 88.50 148.12 52.80 523 ! BI993724 1031003D12.x2 C. re
gb_est2:BI993724 + 88.50 147.65 56.12 546 ! BI993724 1031003D12.x2 C. re
gb_gss:AQ345896 - 88.50 147.29 58.76 564 ! AQ345896 RPC111-126123.TU RP

seq_name: gb_gss:AG078683

seq_documentation_block:
LOCUS AG078683 645 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-074B02.F, genomic survey sequence.
ACCESSION AG078683
VERSION AG078683.1 GI:16630485
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone: PTB-074B02.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 645)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
LIBRARY Sequencing: -21M13
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
source 1. 645
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-074B02.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 104 a 152 c 213 g 176 t
ORIGIN
alignment_scores:
Quality: 110.50 Length: 136
Ratio: 1.811 Gaps: 5
Percent Similarity: 44.853 Percent Identity: 26.471
alignment_block:
US-09-674-779-2 x AG078683/rev ..
Align seg 1/1 to reverse of: AG078683 from: 1 to: 645
24 ILeProThrasnProGlnValSerProIle..... 33
:::|||||.....
561 TTGCCAACAAACCCCAATCCCTCATTTGGCGCTGTGGCAACCGTTGACG 512
34LysThrProSerValLeu..... 39
511 CACTTTTGGCTGTACAAACCCCTCAGATAGTGGAAAGAACACACAG 462
40IleThrLysAsp 43

```

461 AGACTGAGGACGAGAGAAAGAGAGACGGGAGCCACACAAAGAC 412
      ::::::::::::::
44  LysIleGlyAspHisHisThrHisGluHisAspGluSerValSerHisVa 60
      ::::::::::::::
411 ACCTACATACACACACACACACACACACACACACACACACACATC 362
      ::::::::::::::
60  lclyleuclnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaThrL 77
      ::::::::::::::
361 AGGGTCATACAGCAGACAGAG.....CAGCCAGGGTCA 327
      ::::::::::::::
77  ysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsn 93
      ::::::::::::::
326 CGCAGAGCTCCCGTCGATCAGATCAGATGGGACGACTTGGGG 277
      ::::::::::::::
94  TyrLeuProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnA 110
      ::::::::::::::
276 ACACCTCACTCCA.....ACCAACACCATCCGTCGAGG 245
      ::::::::::::::
110 laCysGlyHisclupProTyrGlnLeuProGluHisLeuTrpGlyGln 126
      ::::::::::::::
244 CCTCAGGCTGTATCCCTATCTGTACCCCGCTGCGCGCTGGGTCC 195
      ::::::::::::::
127 IleVal 128
      ::::::::::::::
194 ACCATC 189
      ::::::::::::::

seq_name: gb_uss:AZ638282

seq_documentation_block:
LOCUS      AZ638282                640 bp    DNA        linear    GSS 13-DEC-2000
DEFINITION  lM0498P02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0498P02 F, DNA sequence.
ACCESSION   AZ638282
VERSION     AZ638282.1  GI:11760472
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1 (bases 1 to 640)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: dgunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0498 row: P column: 02
            Seq primer: CGTGTAAACGAGCGGCGAGT
            Class: plasmid ends
            High quality sequence stop: 640.
FEATURES    Location/Qualifiers
            1..640
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0498P02"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource

```

```

(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (gi14732114.gb/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      181 a      181 c      122 g      155 t      1 others
ORIGIN

alignment_scores:
    Quality: 105.50      Length: 150
    Ratio: 1.353      Gaps: 7
Percent Similarity: 52.000      Percent Identity: 28.667

alignment_block:
US-09-674-779-2 x AZ638282 ..
Align seg 1/1 to: AZ638282 from: 1 to: 640

35  ThrProSerValLeuIleThrLysAspLysIleGlyAspHis..... 48
    |||||:::||||:::|||||
34  ACACCTAAGGTTGTACTCTGTACCTGTACACACACACACACACACAC 83
    |||||:::|||||
49  HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaH 65
    |||||:::|||||
84  ACACACACACACACACACACACACACACGACGACACTGTGTGTACACAC 133
    |||||:::|||||
65  isPheGluThrTrpLeuGln.MethHisHisAlaThrLysGlnGluValVa 81
    |||||:::|||||
134 ACATACAACTACTGTCCAGGAAGAATAAAACCAAGCAAGGTTCTTCT 183
    |||||:::|||||
81  lArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPROM 98
    |||||:::|||||
184 GAAA.....CCCTCGCAAGT.....CCACAAA 206
    |||||:::|||||
98  etSerGlnLeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGlu 114
    |||||:::|||||
207 ACCAGGCTGCTTAAGTACT.....CCA 229
    |||||:::|||||
115 ProTyrGlnLeuProGluHisLeuTrpGlyGlnIleValProThrLe 131
    |||||:::|||||
230 TCACACTCTCTAGCCCTTGAGCACACATGATGAAGATCCTCAGGCACAT 279
    |||||:::|||||
131 uHisLeuTyrGlnAspLeuLysSerArgGlyIle.....LeuProAlaA 146
    |||||:::|||||
280 TCACCTTCATCTTGTATCTTCTCAGCCCTTGGATTACTATGCTTCTCTAT 329
    |||||:::|||||
146 snThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAla 162
    |||||:::|||||
330 TAGTCTCCATCCATAGTCTCTACCGGGGCGGACGTCCTCATCTCATGCACA 379
    |||||:::|||||
163 Gly.....GlyAlaAlaMetSerLysHisLeuThrAsnSer 174
    |||||:::|||||
380 GGAGAGGTCCTTAATGGGCAGTTAATGACCCCTCTTCTGCTGGGTCA 427
    |||||:::|||||

seq_name: gb_uss:AZ953717

seq_documentation_block:
LOCUS      AZ953717                632 bp    DNA        linear    GSS 27-APR-2001
DEFINITION  2M0219P02F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
            clone UUGC2M0219P02 F, DNA sequence.
ACCESSION   AZ953717

```

1

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
TITLE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1270)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cyapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM12319 row: j column: 11
High quality sequence stop: 469.
Location/Qualifiers
1. .1270
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5573098"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 272 a 475 c 242 g 281 t
ORIGIN
alignment_scores:
Quality: 98.50 Length: 206
Ratio: 1.048 Gaps: 6
Percent Similarity: 45.631 Percent Identity: 24.757
alignment_block:
US-09-674-779-2 x BM467307 ..
Align seg 1/1 to: BM467307 from: 1 to: 1270
13 ILeSerSerMetLeuValAlaCysSerAlaProIleProThrAsnProgl 29
744 ATCTCTCCATAACTCCCATCGCTCTCCCATCAGTCCTACCCACCACC 793
29 nValSerProIleLysThrProSerValLeuIleThrLysAspLysIle 46
794 TCAACATCCAGTAAATCTCCC.....G 816
46 lYAspHisThrHisGluHisAspGluSerValSerHisValGlyLeu 62
817 GCTACACATGTCAT..... 833
63 GlnAlaHisPheGluThrTrpLeuGlnMetHisAlaThrLysGlnGl 79
834ACACCAATTC 844
79 uValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr.Leu 95
845 ACTCACCGCTCCGCCCATATTTGGGTCCAGATCTCTCATACCTCA 894
96 ProProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGl 112
895 CCGCCGGCTCTCGAACTCTCGTGTATCAGACGAACTGCTCTCCTC 944
112 yHisGluProTyrGlnLeuProGluHisLeuTrpGlyGlnIleValP 129

945 CCGCTTCTCTTCTGCTGCTCCCTCCCTCTTTGTGGAATCGCTACTC 994
 129 roThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAla 145
 ||||| : : : ||| : : : |||
 995 CGTCTCCACCACTGAAGTGCATGTCAGCAGTGTGCATATATCCGCGC 1044
 146 AsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAl 162
 : : : ||||| : : : |||
 1045 CCCCAACACATCGCTCA...TATCGACATGTACCCCTC..... 1080
 162 aGlyGlyAlaAlaMetSerLysHisLeuThrAsn...SerAlaIleAspI 178
 ||||| : : : ||| : : : |||
 1081TCTCAATCACACACACACACGACCAATCTCCCTATTATCTCC 1123
 178 leTrpValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGln 194
 : : : ||||| : : : |||
 1124 CTCCTTCCCATCTCGATGTTGTAATCGCGTCCAAAGCACTCGCTCAA 1173
 ||||| : : : ||| : : : |||
 195 AsnArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsnPheG1 211
 ||||| : : : ||| : : : |||
 1174 AATCGTCTCTACGCTCTCTCCCGCGTGTTCGCGCCGCCACCCACATCGC 1223
 211 yLeuGlyLeuTyrAla 216
 : : : ||||| |||||
 1224 ATCATCTCTATACGCG 1239

seq_name: gb_gss:AG140443

seq_documentation_block:
 LOCUS AG140443 643 bp DNA linear GSS 04-NOV-2001
 DEFINITION Pan troglodytes DNA, clone: PTB-155D18.F, genomic survey sequence.
 ACCESSION AG140443
 VERSION AG140443.1 GI:16670121
 KEYWORDS GSS: GSS (genome survey sequence)
 SOURCE Pan troglodytes male lymphoblast DNA, clone.lib:PTB Chimpanzee Male
 BAC Library clone:PTB-155D18.F.

ORGANISM

Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 (sites)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE

BAC end sequences of Library PTB

JOURNAL

Unpublished
 2 (bases 1 to 643)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimpanzee@sc.riken.go.jp; URL: http://hgp.gsc.riken.go.jp/
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.

COMMENT

PRIMERS
 Sequencing: -21W13

LIBRARY

Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.

FEATURES

source

1. .643
 Location/Qualifiers
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-155D18.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone.lib="PTB Chimpanzee Male BAC Library"
 143 a 240 c 101 g 158 t 1 others

BASE COUNT

ORIGIN

alignment_scores:
 Quality: 98.00 Length: 208
 Ratio: 0.990 Gaps: 13
 Percent Similarity: 47.596 Percent Identity: 26.923
 alignment_block:
 US-09-674-779-2 x AG140443 ..
 Align seg 1/1 to: AG140443 from: 1 to: 643
 10 ThrThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProTh 26
 ||||| : : : : : |||
 43 AGCAGTTGCCGCCAGNCCTTGATTGCTCTCCAGTACCTCTTATCAAG 92
 : : : : : : : : : : :
 26 rAsnProGlnValSerProIleLysThrProSerValLeuIleThrLysA 43
 : : : : : : : : : : :
 93 ACACCTGCCTACTTCTCCCATCATGGAGCCGTCA.....CACAAAG 133
 : : : : : : : : : : :
 43 sPlysIleGlyAspHis...HisThrHisGluHisAspGluSerValSe 58
 ||||| : : : ||||| |||||
 134 ATACACACTGCACACATAGACACACACACACAC.....AC 171
 : : : : : : : : : : :
 58 rHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisHis 75
 ||||| : : : ||||| |||||
 172 ACACACACACACACACACACACACACACACACCCCTGCCCATGAAG 221
 : : : : : : : : : : :
 75 laThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeu 91
 : : : : : : : : : : :
 222 TTACTCGAAGAACTGTTCCACTTCAA..... 250
 92 GlyAsnTyrLeuPro.....ProMetSerGlnLeu... 101
 ||||| : : : ||||| : : : |||||
 251CTTCCCTTTTCCTTAGTCTTGGTCACCTTTTCTCTCTGCCC 291
 : : : : : : : : : : :
 102 LeuThrThrAlaArgSerTrpGln.....AlaC 111
 ||||| : : : ||||| : : : |||||
 292 ACTGCAACCCCTCCTAGTGTGCGCACCTGCTGAATGCCATAAGTTCTGCT 341
 : : : : : : : : : : :
 111 yGlyHisGluProTyrGlnLeuProGluHisLeuTrpGlyGlnle 127
 ||||| : : : ||||| : : : |||||
 342 GTCACATCCACAGGTCTCTGTCAGTCCGCCACCCAGG.....CAGCCT 382
 : : : : : : : : : : :
 128 ValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuPr 144
 : : : : : : : : : : :
 383 CTTCCTCTCTTGCACCTA.TGGAGTCCCTGACACTCTCGGCTGCTCCC 431
 : : : : : : : : : : :
 144 oAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnC 161
 : : : : : : : : : : :
 432 CATATA.....CTGCGCATGCTGCTCCGAAG..... 456
 : : : : : : : : : : :
 161 ysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAsp 177
 : : : : : : : : : : :
 457GCTGCCCTTCTCTTCATCTC..... 477
 : : : : : : : : : : :
 178 lIeTrpValProAspLeuGluIleLysSerGlnAlaLeuTyr.....G1 192
 ||||| : : : ||||| : : : |||||
 478 ...TGGTATCCATCCATCTCTATCTCCCACTACCACTACTGCTGCTCTC 524
 : : : : : : : : : : :
 192 uLeuGlnAsnArgLeuCysGln 199
 : : : : : : : : : : :
 525 ATTTTCAAAACCATCTGTGTGAA 546

seq_name: gb_est2:BG214490

seq_documentation_block:

LOCUS BG214490

DEFINITION RST34131 Athersys RAGE Library Homo sapiens

ACCESSION BG214490

VERSION BG214490.1 GI:13740511

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

621 bp mRNA linear EST 21-APR-2001


```

87  LeuGlnSer.....ArgLeuGlyAsnTyrLeuProMetSerGlnL 101
|||||..... |||||
478  ATGTGATGAATGCATCCGACTGGGA..... 502
101  euLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGln 117
|||||..... |||||
503  .....CAAATACCGGACTGGCCG..... 523
118  LeuProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTy 134
523  ..... 523
134  rGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgS 151
|||||..... |||||
524  .....AGATCTGGAGGTATACCCAGACCACT..... 553
151  erValTyrArgAsnProGluLeuAsnGlnCysAlaGlyClyAlaMet 167
|||||..... |||||
554  .....CAATGCAAGCGCAACCCCAACACA 577
168  SerLysHisLeuThrAsnSerAlaIleAspIleTyrValPro...AspLe 183
|||||..... |||||
578  GACATGATGATCACCAGCAACAGCGCAAGCATGCGCCCATGACT 627
183  uGlu.....IleLysSerGlnAlaLeuTyrGluLeuG 194
|||||..... |||||
628  CGAGAGCTCCATCGCGCAACCAATATCTCCGAGACACACCCCAACAA 677
194  InAsn...ArgLeuCysGlnTyrTrpLeuGluHisGlyGluAsnGlnAsn 209
|||||..... |||||
678  AAAATGGACGAAGAACAACATACAAAGGAAGAACCGCACACGACACAG 727
210  PheGlyLeuGlyLeuTyrAlaThrGlyAlaIleHisLeu..... 222
|||||..... |||||
728  ...GGAATCTGCTGCAACAGACGACAGATACACACCCCTACTCAAC 774
223  .....AspThrGlnGlyPheArgLysTrpGlyA 232
|||||..... |||||
775  CACACCCATCCCCCAACACGACACACGACGCAACAGG..... 814
232  laGlnPheSerGluThrAsnSerIleCysArgHisValLeuProLysAsn 248
|||||..... |||||
815  .....ACGCATACACAGACGACGACGCGCACCGCCGCCAATAA 850
249  Lys 249
851  CGA 853

```

seq_name: gb_gss:CNS04OFM

```

seq_documentation_block:
LOCUS      CNS04OFM                      813 bp      DNA      linear      GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            124008 of library G from Tetraodon nigroviridis, genomic survey
            sequence.

```

```

ACCESSION  AL299947.1  GI:8039088
VERSION    AL299947.1  GI:8039088
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetraodon nigroviridis.
            Tetraodon nigroviridis.

```

```

ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.

```

```

REFERENCE  1 (bases 1 to 813)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
JOURNAL   unpublished
REFERENCE  2 (bases 1 to 813)
AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

```

```

TITLE      Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
REFERENCE  Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
JOURNAL    Unpublished
            3 (bases 1 to 813)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT    This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.

```

FEATURES

```

source     1..813
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="124008"
            /note="lib="G"
            /note="Genoscope sequence ID : C0BG124BH04SPL-end :
            PUC-Ori"

```

```

BASE COUNT 234 a 191 c 189 g 189 t 10 others
ORIGIN

```

alignment_scores:

```

Quality: 96.00      Length: 61
Ratio: 2.462        Gaps: 0
Percent Similarity: 63.934 Percent Identity: 34.426

```

alignment_block:

US-09-674-779-2 x CNS04OFM/rev ..

Align seg 1/1 to reverse of: CNS04OFM from: 1 to: 813

```

130 ThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyIleLeuProAlaAs 146
|||||..... |||||
209 ACCCTCAGCTCATCCAGAGCTGATTTACTCTCAGATGCTGCCAGAGTC 160
|||||..... |||||
146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGlnCysAlaG 163
|||||..... |||||
159 CTCCTCTCAGCGGCTCTGTAGTTCAAACCCAGACAGATGCTGCTGCTCG 110
|||||..... |||||
163 lyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrp 179
|||||..... |||||
109 GCACCTCTGGGCTCCCAACACACAGGAGATGTGAGTATCTCCCTCTGG 60
180 ValProAspLeuGluIleLysSerGlnAlaLeu 190
|||||..... |||||
59 GTGCCAGACCTGGAGTCTGGATGCATCTT 27

```

seq_name: gb_est2:BE442326

```

seq_documentation_block:
LOCUS      BE442326                      454 bp      mRNA      linear      EST 25-JUL-2000
DEFINITION 925017E08.xl C. reinhardtii CC-2290, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.

```

```

ACCESSION  BE442326
VERSION    BE442326.1  GI:9441841
KEYWORDS   EST.
SOURCE     Chlamydomonas reinhardtii.
            Chlamydomonas reinhardtii.

```

```

ORGANISM   Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadales; Chlamydomonas.

```

```

REFERENCE  1 (bases 1 to 454)
AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
JOURNAL   unpublished (2000)
COMMENT   Contact: Elizabeth H. Harris
            DCMB Box 91000

```

Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.

FEATURES

Location/Qualifiers
1. 454

/organism="Chlamydomonas reinhardtii"
/strain="CC-2290 wild type mt- S1 D2"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-2290, normalized, Lambda Zap II"

/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library was constructed by John Davies and Jeffrey McDermott. RNA was isolated from strain CC-2290 (Minnesota isolate of C. reinhardtii) grown to mid-log phase in TAP (acetate containing) medium in the light. PolyA mRNA was purified, and cDNA was synthesized and directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 143 a 140 c 101 g 70 t
ORIGIN

alignment_scores:
Quality: 95.50 Length: 84
Ratio: 1.910 Gaps: 5
Percent Similarity: 59.524 Percent Identity: 35.714

alignment_block:

US-09-674-779-2 x BE442326 ..
Align seg 1/1 to: BE442326 from: 1 to: 454

60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMetHisHisAlaTh 76
183 ATTGGCAACAGCAGCATTCGACACA.....CAGCAAAAGCGAGCAGC 226

76 rlyGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
227 AAAACAACACGCTATCCGCCAAGACACA.....TAAATCGCGTGCGCA 270

93 snTyrLeuProProMetSerGlnLeuLeuThrAlaArg...SerTrp 108
271 CC.....AATCAAGGACCAAAACTCACAAGAGACGCAAGCTGG 314

109 GlnAlaCysGly.....HisGluProTyrGlnLeuProProGluHisLe 123
315 CAAAACTCGCTCAAGTTTCACAGGCGCAGCGGCGTGCCACATGTGCG 364

123 utrPGlyGlnIleValProThrLeuHisLeuTyrClnAspLeuIysSera 140
365 CTGGGGCAGCGCTATCGGAACAGCAGCTGTACCGTCTGTCGCCGACA 414

140 rg 140
415 GA 416

seq_name: gb_gss:BH032955

seq_documentation_block:

LOCUS BH032955 774 bp DNA linear GSS 17-JUL-2001
DEFINITION RPI-24-294G15-TV RPI-24 Mus musculus genomic clone RPI-24-294G15
DNA sequence.
ACCESSION BH032955
VERSION BH032955.1 GI:14804300
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 774)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M., Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E., Russell, D., de Jong, P., and Fraser, C.M.
Mouse BAC End Sequences from Library RPI-24
Unpublished (1999)
Other_GSSs: RPI-24-294G15.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi1.org/bacpac/orderingframe.htm>). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 294 row: G column: 15
Seq primer: T7
Class: BAC ends.

FEATURES

Location/Qualifiers

Source

1. 774
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPI-24-294G15"
/clone_lib="RPI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; RPI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT 229 a 265 c 85 g 195 t
ORIGIN

alignment_scores:

Quality: 95.00 Length: 76
Ratio: 2.159 Gaps: 1
Percent Similarity: 57.895 Percent Identity: 27.632

alignment_block:

US-09-674-779-2 x BH032955 ..

Align seg 1/1 to: BH032955 from: 1 to: 774

3 AspPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLeuValAl 19
83 AACACCAACATCATCATCACCACTGTTATCGCAATCATCATCAT 132
19 acySSerAlaProIleProThrAsnPro.....GlnValSerP 32
133 AGCGCGCAATTGTCACACCTTCACCCAGTCATCACAAAGTACCATAGCA 182
32 rotleYsThrProSerValLeuIleThrIysAspIysIleGlyAspHis 48
183 CTGTGTCACCAATTCATCATCACCAACAGGACCAACCATCAACAT 232
49 HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAla 65
233 CATCACCAACCAATTCATCATCATCACCAACAGGACCAACCATCA 282
65 sPheGluThrTrpLeuGlnMetHisHis 74
283 TTATATCATCATCATCACCAACAGGACCAACCATCAACATCA 310

seq_name: gb_est2:BF627823

```

seq_documentation_block:
  LOCUS      BF627823      808 bp      mRNA      linear      EST 22-OCT-2001
  DEFINITION  HVSMED0005009f Hordeum vulgare seedling shoot EST library
  HVCDNA0002 (Dehydration stress) Hordeum vulgare cDNA clone
  HVSMED0005009f, mRNA sequence.
  ACCESSION  BF627823
  VERSION    BF627823
  KEYWORDS   EST.
  SOURCE     barley.
  ORGANISM   Hordeum vulgare
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
              ; Triticeae; Hordeum.
  REFERENCE  1 (bases 1 to 808)
              Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
              ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
              ,R.D., Oates,R. and Main,D.
  TITLE      Development of a genetically and physically anchored EST resource
              for barley genomics: Morex drought-stressed seedling shoot cDNA
              library
  JOURNAL    Unpublished (2001)
  COMMENT    On Dec 19, 2000 this sequence version replaced gi:11891981.
              Contact: Wing RA
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Tel: 864 656 7288
              Fax: 864 656 4293
              Email: rwing@clemson.edu
              Total hg bases = 275
              Seq primer: AATTAACCTCTCACTAAAGG
              High quality sequence start: 4
              High quality sequence stop: 753.
              Location/Qualifiers
                1..808
                  /organism="Hordeum vulgare"
                  /cultivar="Morex"
                  /db_xref="taxon:4513"
                  /clone="HVSMED0005009f"
                  /clone_11b="Hordeum vulgare seedling shoot EST library
                  HVCDNA0002 (Dehydration stress)"
                  /tissue_type="Seedling shoot"
                  /lab_host="TJC121"
                  /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
                  Seeds were surface sterilized then germinated under axenic
                  conditions in the dark at room temperature on filter paper
                  with water, nystatin and cefotaxime in covered
                  crystallization dishes. Five-day old seedlings were
                  incubated at 90% RH for 24 hr. Shoots were then harvested,
                  total RNA was prepared, poly(A) RNA was purified, one
                  primary unamplified cDNA library was made, 600000 pfu were
                  in vivo excised to give pluescript SK(-) cDNA phagemids.
                  These steps were performed in the TJ Close laboratory at
                  the University of California, Riverside (Choi, Close,
                  Fenton). Phagemids were plated and picked at the Clemson
                  University Genomics Institute (CUGI) (Begum, Palmer,
                  Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                  sequencing and sequence analysis were performed at CUGI
                  (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                  The sequence has been trimmed to remove vector sequence
                  and contains a minimum of 100 bases of phred value 20 or
                  above. For more details on library preparation and
                  sequence analysis see
                  http://www.genome.clemson.edu/projects/barley. To order
                  this clone see http://www.genome.clemson.edu/orders
                  Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                  Genetically and physically anchored EST resources for
                  barley genomics. Barley Genetics Newsletter 31:29-30.
                  (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)."
  BASE COUNT 101 a 256 c 321 g 130 t
  ORIGIN

```

```

alignment_scores:
  Quality: 95.00      Length: 177
  Ratio: 1.159      Gaps: 7
  Percent Similarity: 46.328      Percent Identity: 22.034

alignment_block:
  US-09-674-779-2 x BF627823/rev ..
  Align seg 1/1 to reverse of: BF627823 from: 1 to: 808

19 AlaCysSerAlaProIleProThrAsnProGlnValSerProIleLysTh 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 GCCTGGCGCCAGCCCTACCTACC...CCCCAGCGCAGCCCGCGCGCAG 476

35 rProSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisG 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
475 CCCCACGCCCCCTC.....CGCTTCGGTGCCATTCCCGTCCC 439

52 LuHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluThr 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 .....CGCTTCGGTGCCATTCCCGTCCC 439

69 TrpLeuGlnMetHis.....HisAl 75
|||||:|||||:|||||:|||||:|||||:
438 TCCTCGCCGACGACGCGCGGCGAGCGCGGCGCGCGCGCGCGCACCT 389

75 aThrLysGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuG 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 CCGCGCAGCGCAGCGCGCTGCGCTTCCAGTAGTACTCCCCCGCGCAGC 339

92 Ly.....AsnTyrLeuProProMetSerGlnLeuLeu 102
|||||:|||||:|||||:|||||:|||||:
338 GCTAGCGCGGTGGCTAGTACAGCAGGTAGCCCCCGCATGCAG..... 298

103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
297 TTGTGACGCGCAGGTGGAGCGGTGTGCGCGCGCGCGCGCGCGCGCGC 248

119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
|||||:|||||:|||||:|||||:|||||:|||||:
247 CCGCGCGCGGAGC.....ACCGCGCTGAGGAGATCTCGTTGTACCACT 204

136 sPLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerVal 152
|||||:|||||:|||||:|||||:|||||:|||||:
203 CCGTCTCTCTCCCG..... 190

153 TyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLy 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 ...AACAGCGCCAGTAATCTCTCCCGCGCGCGCGCGCGCGCGCGCGC 143

169 sHisLeuThrAsnSerAlaIleAspIleTrp 179
|||||:|||||:|||||:|||||:|||||:|||||:
142 CGCATGTGCGCGCGCGGTGGTGGTGGTGGTGG 112

seq_name: gb_est2:BE974669

seq_documentation_block:
  LOCUS      BE974669      1044 bp      mRNA      linear      EST 04-OCT-2000
  DEFINITION  60168070782 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:3951145 3',
              mRNA sequence.
  ACCESSION  BE974669
  VERSION    BE974669
  KEYWORDS   EST.
  SOURCE     human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE  1 (bases 1 to 1044)
              NIH-MGC http://mgc.nci.nih.gov/.
  AUTHORS    National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE      Unpublished (1999)
  JOURNAL
  COMMENT    Contact: Robert Strausberg, Ph.D.

```

```

152 altyrArGAsnProGluLeuAsn...GlnCysAlaGlyAlaAlaMet 167
      :::::||||| :::: ||| |||||
579 CTTgCGGTGCTCTGCGGGCAATGAAACACCCGCGGCGCTCCGA 530
      :::::||||| :::: ||| |||||
168 SerLysHis.....LeuThrAsnSer..... 174
      :::::||||| |||||
529 GGGCGGCCCCGTCGGTCAAAACAACCTCGTGAATCAGGGGTCGCACN 480
175 .AlaIleAspIleTrpValProAspLeuGluIleLysSerGlnAlaLeuT 191
      |||||::: :::::
479 CAGATAGATGTT.....GTGCACACACACACAGAG 448
191 yrGluLeuGlnAsnArg.....LeuCysGln 199
      |||||:::|||||
447 AGCACCTTATGAGTCGAGGAAGATTATGTGCAA 416

```

seq_name: gb_est2:BI728822

seq_documentation_block:
LOCUS BI728822 599 bp mRNA linear EST 19-SEP-2001
DEFINITION 103110IG02.y1 C. reinhardtii CC-1690, Stress II (normalized),
Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI728822
VERSION BI728822.1 GI:15704517
KEYWORDS EST.

REFERENCE
 1. (bases 1 to 599)
 AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrager, J., Sillflow, C. and Stern, D.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031
 JOURNAL Unpublished (2001)

JOURNAL COMMENT

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

1. 599
/location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/Clone.lib="C. reinhardtii CC-1690, Stress II (normalized)
, Lambda Zap II"
/note="vector: pBluescript II SK-; Site.1: EcoRI; Site.2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Wells et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda Zap clones by superinfection with ExAssist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."

114 a	186 c	204 g	95 t
-------	-------	-------	------

Quality:	94.50	Length:	184
Ratio:	1.062	Gaps:	10

ref: 010-234-0970

ref: 010-234-0970


```

37 .....ServValLeuThrLysAspLysIleG 46
269 AGGCAGCCACCGCGCTGGTATCCATCGCATCCCAATACAGGTA 318
46 lyAspHis...HisThrHisGluHisAspGluSerValSerHisValGly 61
319 TAGAGCACAGCTCCACTAGTAGGCAC.....CACATAGCC 353
62 LeuGlnAla.....HisPheGluThrTrpLeuGlnMetHisHisAlaLath 76
354 CGCTAGCCCTGGTGCATCTACCATCCGCACTTCAG...CACAGAACTGC 400
76 rlyGlnGlnValValArgTrpGlnAlaTyrLeuGlnSerArgLeuGlyA 93
401 CCAACCAAGCATCGTACCACAGCAGCGGAGCCGACAGGCCCCCAATCAC 450
93 snTyrLeuPro.....ProMetSerGln.....LeuLeu 102
451 CATGGACTCTCTCGAACTGCAGCGCCGAGCATCAACAACCTCAGGACACCAT 500
103 ThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPr 119
501 ACCGCGCTACATACAGAAATGCACAGGACACCGCGCCCAACACACACC 550
119 oProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnA 136
551 CAGCAGACACCTTAAGCAGCTCGCACCCCGC.....CCACTACAACAC 597
136 spLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerVal 152
598 CCAGCGCAACTATGGCCGCTGTCTAATGGCACTAGGTTGATATCCACG 647
153 Tyr.....ArgAsnProGluLeuAsnGlnCysAlaG1 163
648 TACCCATCAAGCGCCCTTCCTGGTGGCTTCACCTCCACGCGCGTGGCTG 697
163 yGlyAlaAlaMetSerLysHisLeuThrAsnSer..... 174
698 CGCAATGGCACACCGACCCCATGGACACATTCACCCCATGGCTCTACCA 747
175 .....AlaIleAspIleTrpValPro 181
748 CTACACCCATGTGATTGTGGCTCCG 773

```

seq_name: gb_est2:BE426799

```

seq_documentation_block:
LOCUS BE426799 631 bp mRNA linear EST 24-JUL-2000
DEFINITION WHE0332_C08_F16ZS Wheat unstressed seedling shoot cDNA library
Triticum aestivum cDNA clone WHE0332_C08_F16, mRNA sequence.
ACCESSION BE426799
VERSION BE426799.1 GI:9424642
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 631)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
Seaton, C.L. and Tong, J.C.
The structure and function of the expressed portion of the wheat
genomes
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov

```

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
seq primer: Stratagene SK primer.

FEATURES

```

source
1..631
Location/Qualifiers
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0332_C08_F16"
/tissue_type="Etiolated shoot"
/dev_stage="Five day old seedling"
/lab_host="E. coli SOLR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Shoots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared, a cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
BASE COUNT 126 a 213 c 163 g 129 t
ORIGIN

```

alignment_scores:

```

Quality: 92.00 Length: 220
Ratio: 1.122 Gaps: 11
Percent Similarity: 37.273 Percent Identity: 22.727

```

alignment_block:

```

US-09-674-779-2 x BE426799 ..
Align seg 1/1 to: BE426799 from: 1 to: 631
23 ProIleProThrAsnProGlnValSerProIleLysThrProSerVal 39
|||||:|||||:|||||:|||||:|||||:
6 CCCTGCTGCTGCCAGGCCCAATCTCCCT.....CCCTCTGTGCG 46
39 uIleThrLysAspLysIleGlyAspHisThrHisGluHisAspGluS 56
|||:|||||:|||||:|||||:|||||:
47 TCTC.....GAATCTCCCGCA 63
56 erValSerHisValGlyLeuGln.....AlaHisPhe 66
||:|||||:|||||:|||||:|||||:
64 GCATCGACCATGTCGCGCTACTGCGGAAGTACAAGGATGAGCTCATCA 113
|||||:|||||:|||||:|||||:|||||:
67 GluThrTrpLeuGlnMetHis.....HisAlaThrLysGlnG1 79
|||:|||||:|||||:|||||:|||||:
114 GAACGCTGCCTACATTGGCACCCCTGGCAAGGTATCTCTGCTGCCGACG 163
79 uValValArgTyr.....GlnAlaTyrLeuGlnSerArg..... 90
|||:|||||:|||||:|||||:|||||:
164 ACTCACCGGCAACCATCGGCAAGCGCTTCGCGAGCTCTCTTCGCACCCCTGG 263
91 .....LeuGlyAsnTyr 94
|||||:|||||:|||||:|||||:
214 GTTGAGGACAACCGTCGCGCCCTCGTGAGCTCTCTTCGCACCCCTGG 263
95 LeuProProMetSerGlnLeuLeuThrThrAlaArg..... 106
|||||:|||||:|||||:|||||:
264 AGCCCTCCAGTACCTCAGCGCGGTGATCTCTTCGAGGAGACCCCTGTACC 313
107 .....SerTrpGlnAlaCysGlyHisGluProTyr.....GlnL 118
|||||:|||||:|||||:|||||:
314 AGAGCACCAGGGTGGCAAGCCCTTCGTCGACATCTCAAGGAGGGCAAC 363
118 euProProGluHisLeuTrpGln..... 126
|||||:|||||:|||||:|||||:

```

364 GTCTCCTCGCATCAAGTGGACAAGGTACCGTCGAGCTCGCTGGAAC 413
 126 126
 414 CAACGGTGAGACCACTACCCAGGGCTTTGATGACCTTTGGCAAGCGGTGTG 463
 127 127
 464 CCAAGTACTATGAGCTGGTGGCCCGCTTCCACAGTGGCGCTGCTGCCCTC 513
 135 lnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSer 151
 151 151
 514 AAGATGGCTCCACGAGGCATC.....ACA 539
 152 ValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSe 168
 168 168
 540 GCTTTCATCGACCAAA.....TGCTCAGGGCTGGCTGCTGTATG 580
 168 rLysHisLeu 171
 171 171
 581 CCATCATCTG 590

seq_name: gb_estl:BB821040

seq_documentation_block:

LOCUS BB821040 440 bp mRNA linear EST 19-NOV-2001
 DEFINITION BB821040 RIKEN full-length enriched, mammary gland RCB-0526
 Jyg-MC(A) cDNA Mus musculus cDNA clone G830010A14 3', mRNA
 sequence.

ACCESSION BB821040

VERSION BB821040.1 GI:16993669

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae; Mus.

AUTHORS 1 (bases 1 to 440)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
 Hayatsu, N., Hiramoto, K., Hirozane, T., Imotani, K., Ishii
 Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
 Nakamura, M., Nishi, K., Nomura, K., Okazaki, Y., Okido, T.,
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
 Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
 A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, H., Toya, T.,
 Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
 2001).

TITLE Unpublished (2001)

JOURNAL

COMMENT

Contact: Yoshinide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.

FEATURES Location/Qualifiers

source

1. 440
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RIKEN full-length enriched, mammary gland
 RCB-0526 Jyg-MC(A) cDNA"
 /tissue_type="mammary gland"
 /cell_line="RCB-0526 Jyg-MC(A)"
 124 a 132 c 98 g 86 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 91.50 Length: 184
 Ratio: 1.204 Gaps: 5
 Percent Similarity: 41.304 Percent Identity: 20.652

alignment_block:

US-09-674-779-2 x BB821040 ..

Align seg 1/1 to: BB821040 from: 1 to: 440

11 ThrLeuIleSerSerMetLeuValAlaCysSerAlaProIleProThrAs 27
 26 ACCATCATATCACCTGCTGCTCAAGGCTGAGCTTCAGAGCCCTACTCA 75
 27 nProGlnValSer.....ProIleLysThrProSerVal.....L 39
 76 GCCCAGCAGCTGCTGCCACCCGCTACCCAGTCCACCTCCCGCCGAGG 125
 39 euIleThrLysAspLysIleGlyAspHisHisThrHisGluHisAspGlu 55
 126 CCTACAGCCAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175
 56 SerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMe 72
 176 ACATCTGGCTAC..... 187
 72 ThisHisAlaThrLysGlnGluValValArgTyrClnAlaTyrLeuGlnS 89
 188TACTATGACC 197
 89 erArgLeuGlyAsnTyrLeuProPrometSerGlnLeuLeuThrAla 105
 198 CCAGACTGGTCTATCATGATCCCAATTCAGATGATGATGATGATGATG 247
 106 ArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuProGluHi 122
 248 CAGAGTCACCAG..... 259
 122 sleuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeuLysS 139
 260TACTATATGGGATGGGAGC 281
 139 erArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyrArgAsn 155
 282 GCGGACCTACATACCTGCTGGAGCAGTCTGCTGATGGACATAAGGAC 331
 156 ProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuTh 172
 332 ACGGGGGCATCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 381
 172 rAsnSerAla.....IleAspIleTrpValProAspL 183
 382 AAGAGACGCCCAACAGATGCCAAGGACATGGATCGGTGGGACCCAGTC 431
 183 eu 183
 ||

seq_name: qb_est2:BG848755

```

498 CAGATGACCGCACACCGCCACTGTGTGGCCGC.CGGCGCTGGAGAG 546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 nMethHisAlaThrLysGlnGluValValArgTyrGlnAlaTyrLeu 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
547 CCTGGATACCCAGCGCGTCCACACCG..... 574
88 InSerArgLeuGlyAsnTyrLeuProMetSerGlnLeuLeuThr 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
575 .....CCGATGTAGTCCCGCTCTGCCACTGTGTGCGAGGAACA 613
105 AlaArgSerTrpGlnAlaCysGlyHisGluProTyrGlnLeuPro 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 GCTGCAGGGTGGCTGGTGGCGC...GCACCGTGCAATGTAGGCCA 658

```

seq_name: gb_gss:CNS01Y2Y

seq_documentation_block:
LOCUS CNS01Y2Y 886 bp DNA linear GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
217L13 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL172483
VERSION AL172483.1 GI:7810540
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL
REFERENCE 2 (bases 1 to 886)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL
REFERENCE 3 (bases 1 to 886)
AUTHORS Direct Submission
TITLE Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL This sequence is a single read and was generated as part of a large
COMMENT scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
Source
1..886
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="217L13"
/clone_lib="g"
/note="Genoscope sequence ID : COAG217CF07LPL1-end : T7"

BASE COUNT 218 a 206 c 226 g 235 t 1 others

alignment_scores:
Quality: 91.00 Length: 90
Ratio: 1.685 Gaps: 5
Percent Similarity: 60.000 Percent Identity: 32.222

alignment_block:
US-09-674-779-2 x CNS01Y2Y/rev ..

Align seg 1/1 to reverse of: CNS01Y2Y from: 1 to: 886

```

6 GlnTyrPheIleThrThrLeuIleSerSerMetLeu.....ValAl 19
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 GAATACACCGCTCTGCGCCCTCCAGCCCTCTGTATTACAGCAGCGGTGC 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19 acySerAlaProIleProThrAsnProGlnValSerProIleLysThr 36
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
399 GCATACATACCCCGCAGCATCCGAGGCGCAGGAACCTCCCTC...GTCC 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
36 roSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGlu 52
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
352 CAGTCTCTAGTTAACAAGAAAGCCCGCGATCATCATGAACACCCA 303
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
53 HisAspGlu.....SerValSerHisValGlyLeuGlnAlaHisPheG 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
302 CACCATGAGGTTAATGACGTTTCAGCACGTT...GTTGAAGCCAC..... 261
67 uThrTrpLeuGlnMethHisHisAlaThrLysGlnGluValValArgTyr 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 .....AGAGTTCTTCTCCAGAGCCAGCATCACGTCACGTCTAC 221
84 InAlaTyrLeuGlnSerArg 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
220 AGCGATATCAGCAGAGCCAG 201

```

seq_name: gb_est2:BE779469

seq_documentation_block:

LOCUS BE779469 940 bp mRNA linear EST 20-OCT-2000
DEFINITION 601464207F1 NTH_MGC_67 Homo sapiens cDNA clone IMAGE:3867816 5',
mRNA sequence.

ACCESSION BE779469
VERSION BE779469.1 GI:10200667
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 940)

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Cloning Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>
Plate: LLAM9614 row: m column: 01

High quality sequence stop: 639.

Location/Qualifiers

1..940

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3867816"

/clone_lib="NIH_MGC_67"

/tissue_type="retinoblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

BASE COUNT 214 a 267 c 276 g 182 t 1 others

ORIGIN

alignment_scores:
Quality: 91.00 Length: 250
Ratio: 0.812 Gaps: 13
Percent Similarity: 44.800 Percent Identity: 22.400

alignment_block:

US-09-674-779-2 x BE779469/rev ..

Align seg 1/1 to reverse of: BE779469 from: 1 to: 940

```

20 CysSerAlaProIleProThrAsnProGlnValSerProIleLysThrPr 36
774 TGTGGGCGACCATTCGCCCTTACCTATCAGTATTATTAGTTGGCCCCCA 725
36 oSerValLeuIleThrLysAspLysIleGlyAsp...HisHisThrHisG 52
724 RAATCGAGGTTTCGCCGAGATCATAGTGCCCTCATGCTCATCCCATG 675
52 LuHisAspGluSerValSerHisValGlyLeuGlnAlaHisPhe..... 66
674 AAGGCCCATCAGGCTGCCAGGCTGCTCTGTGTGTGAGATGAGATTT 625
67 .....GluThrTripleuGlnMetHisHisAlaLath 76
624 TAGGGCTCACTACAGCTGTGGAACCTGGCGCGGTAAATAGAGCAAT 575
76 rlyGlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyA 93
574 CCAGCTTGAC..... 565
93 snTyrLeuProProMetSerGlnLeuLeuThrAlaArgSerTrpGln 109
564 ..TTCTTGCCA.....TAATCACTGAGAGACGTTGCCATG 532
110 AlaCysGlyHisGluProTyrGlnLeuProGluHisLeuTrp..... 124
531 ACACGGAGGTGAACCCAGACACAGTCCGCCACCACCAAGCTGTGGAAC 482
125 .....GlyGlnIleValProThrL 131
481 CACAGAAGCCCTGGAGACCCGTGCATGTGGCGGAGCTTGGCAATCGG 432
131 eu.HisLeu...TyrGlnAspLeuLysSerArgGlyIleLeuProAlaAs 146
431 TCCACAGAGCTCNCATGATCTCTTCCCAATGGTGTAGTGCCTCGG 382
146 nThrGlnIleArgSerValTyrArgAsnProGluLeuAsnGln.CysAla 162
381 CATAGTATTGGCAGCATCTCTTT.....GCCTGTGATAAGTGTCTCA 338
163 GlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTr 179
337 GGGTGGAGAGCTGGGTAGGTGGCCAGTCCGCACTTCATCATGACTGTG 288
179 pValProAspLeuGluIleLysSerGlnAlaLeuTyrGluLeuGlnAsnA 196
287 GGTTCAGTCTACAAACACTGCCCGGGCA.....C 256
196 rglLeuCysGln.....TyrTrp 201
255 ATGCTTGGCAGCCCGCTCTCACTGAAGAAGGTGTGAAGAAATCATCTC 206
202 LeuGluHisGly...GluAsnGlnAsnPheGlyLeuGlyLeuTyrAla... 216
205 CCCCCCATGGTCTGTCTACTTGGCATCTGGCCATCGGCTGGATGGCGT 156
217 .....ThrGlyAlaIleHisLeuAspThr 224
155 GTTCCAGGCACTAGAGCTCCAGGAGGATTTGCCCATCTGGACACC 110

```

seq_name: gb_est2:BJ111789

seq_documentation_block:

```

LOCUS      BJ111789                651 bp      mRNA      linear      EST 23-JAN-2002
DEFINITION BJ111789 unpublished oligo-capped cDNA library, C. elegans L1 stage
            Caenorhabditis elegans cDNA clone yk1147d06 5', mRNA sequence.
ACCESSION  BJ111789

```

BJ111789.1 GI:18271832

EST.

SOURCE

ORGANISM

Caenorhabditis elegans.
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE

1 (bases 1 to 651)

Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.

and Sugano, S.

A complementary view of the C.elegans genome

Unpublished (2002)

Contact: Tadasi Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1. 651

source

/organism="Caenorhabditis elegans"

/strain="N2"

/db_xref="taxon:6239"

/clone_lib="unpublished oligo-capped cDNA library, C.

elegans L1 stage"

/sex="hermaphrodite"

/tissue_type="whole animal"

/dev_stage="L1"

BASE COUNT 174 a 157 c 127 g 193 t

ORIGIN

alignment_scores:

Quality: 90.50 Length: 216

Ratio: 0.887 Gaps: 13

Percent Simlarity: 47.222 Percent Identity: 24.074

alignment_block:

US-09-674-779-2 x BJ111789 ..

Align seg 1/1 to: BJ111789 from: 1 to: 651

```

2 LysAsnPheAsnGlnTyrPheIleThrLeuLeuSerSerMetLeuVa 18
101 GAACAATGGTCAAGATATTCTTTGTCTATGCTCTTTACCAACTTATCAT 150
18 lAlaCysSerAlaProIleProThrAsnProGlnValSerProIleLysT 35
151 ATCTTTGCTCACT.....TCTTCTACGGAGA 176
35 hrProSerValLeuIleThrLysAspLysIleGlyAspHisHisThrHis 51
177 CACCATCTTTGGAATCAAGACAGGACTCTATG.....CACCACCTCTCG 220
52 GluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheGluTh 68
221 CTCACCATCATCAACATTTCA.....CT 243
68 rTrpLeuGlnMetHis.....HisAlaThrLysGlnGluV 80
244 TTGGTTTTCATGCAITTCGTTATGATGACTGTCTTCAATGAGATCAATGC 293
80 alValArgTyr..... 83
294 TCGTAGGTTTCACGGAGCGGTAAACGTGTTTAAAGGATTGGCATCTAAC 343
84 .....GlnAlaTyrLeuGlnSerArgLe 91
344 GAGTTTCTGTGCTATTGGTTTACACATTCATTCCTCAAAATCATATT 393
91 uGlyAsnTyrLeuProProMetSerGlnLeuLeu...ThrThrAlaArgSe 107

```

```

BASE COUNT      147 a   223 c   195 g   135 t
ORIGIN
alignment_scores:
  Quality:      90.00      Length:      106
  Ratio:        1.607      Gaps:        5
  Percent Similarity: 52.830  Percent Identity: 29.245
alignment_block:
  US-09-674-779-2 x BB665280/rev ..
  Align seg 1/1 to reverse of: BB665280 from: 1 to: 700
78  GlnGluValValArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyr 94
   :::::  |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
237 AGGGAAGCCAGAGGATGAAGCA...GTGACAAGCATCGTGGAGGGGC 191
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
94  rLeuProMetSerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaC 111
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
190 CTGGGCCCCAGAGAGGCGTC.....CAGTTCT 162
111 ysglyHis.....GluProTyrGlnLeuProGluHisLeuTrpGly 125
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
161 GTGGACACCTCGCGGCCCTTCGGGTTCCCTCCTCCTACTTG..... 118
126 GlnIleValProThrLeuHisLeuTyrGlnAspLeuLysSerArgGlyTl 142
118 ..... 118
142 eLeuProAlaAsnThrGlnIleArgSerValTyrArgAsnProGluLeuA 159
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
117 .CTGCACGACCGCCCGCCAGCGGAGGACCTGGGATGGCCCGAGCCCTCT 69
159 snGlnCysAlaGlyGlyAlaAlaMetSerLysHisLeuThrAsnSerAla 175
   ::  :::::  |||  :::::  |||  :::::  |||  :::::  |||
68 CA...GCAGGTGGCGGTGCAGCCCTTGGGACGGTTGGCATTCTTGACCT 22
176 IleAspIleTrpValPro 181
   |||  |||  :::::  |||
21 TCCGACAGATGGGTACTCT 4
seq_name: gb_est2:BG368308

seq_documentation_block:
  BG368308 746 bp mRNA linear EST 22-OCT-2001
LOCUS HVSME1001J713f Hordeum vulgare 20 DAP spike EST library HVCDA0010
DEFINITION (20 DAP) Hordeum vulgare cDNA clone HVSME1001J713f, mRNA sequence.
ACCESSION BG368308
VERSION BG368308.2 GI:163225190
KEYWORDS EST.
SOURCE barley.
  ORGANISM
    Hordeum vulgare
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
    ; Triticeae; Hordeum.
REFERENCE
  1 (bases 1 to 746)
  Wing, R., Close, T.J., Kleinohs, A., Wise, R., Begum, D., Frisch, D., Yu
  , Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
  , R.D., Close, S.J., Oates, R. and Main, D.
  Development of a genetically and physically anchored EST resource
  for barley genomics: Morex 20 DAP spike cDNA library
  Unpublished (2001)
  On Mar 8, 2001 this sequence version replaced gi:13257409.
  Contact: Wing RA
  Clemson University Genomics Institute
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Total hg bases = 438
  Seq primer: AATTAACTCTCACTAAAGG

```

```

FEATURES
  Location/Qualifiers
    1..746
    /organism="Hordeum vulgare"
    /cultivar="Morex"
    /db_xref="taxon:4513"
    /clone="HVSME1001J713f"
    /clone_lib="Hordeum vulgare 20 DAP spike EST library
    HVCDA0010 (20 DAP)"
    /tissue_type="20 DAP spike"
    /lab_host="SOLR"
    /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
    plants were grown in the greenhouse at the University of
    California, Riverside (Fenton, SJ Close, TJ Close). Whole
    spikes with awns trimmed were collected at 20 DAP (Fenton
    ). Total RNA was prepared, poly(A) RNA was purified, one
    pfu were in vivo excised to give phuescript SK(-) cDNA
    primary unamplified cDNA library was made, and 1 million
    phagemids in the TJ Close lab at the University of
    California, Riverside (Choi). Phagemids were plated and
    picked at the Clemson University Genomics Institute (CUGI)
    (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
    preparations, DNA sequencing and sequence analysis were
    performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
    , Rambo, Main). The sequence has been trimmed to remove
    phred value 20 or above. For more details on library
    preparation and sequence analysis see
    http://www.genome.clemson.edu/projects/barley. To order
    this clone see http://www.genome.clemson.edu/orders/Also
    see Close TJ, Wing R, Kleinohs A, Wise R (2001)
    Genetically and physically anchored EST resources for
    barley genomics. Barley Genetics Newsletter 31:29-30.
    (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)."
    BASE COUNT      162 a   251 c   189 g   143 t   1 others
    ORIGIN
alignment_scores:
  Quality:      90.00      Length:      214
  Ratio:        1.098      Gaps:        11
  Percent Similarity: 38.318  Percent Identity: 22.430
alignment_block:
  US-09-674-779-2 x BG368308 ..
  Align seg 1/1 to: BG368308 from: 1 to: 746
56 SerValSerHisValGlyLeuGln.....AlaHisPh 66
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
49 AGCATCGACCATGTGGCCTACTCGCGAAGGTACAAGGATGAGCTCATCA 98
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
66 eGluThrTrpLeuGlnMethHis.....LeuGlyAsnTyr 73
   |||  |||  :::::  |||
99 AGACGCTGCCTACATTTGGCACCCCTGGAAAGGTATCTTGTCTGTGAC 148
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
74 .....HisAlaThrLysGlnValValArgTyrGlnAlaTyrLeuGln 88
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
149 GAGTCCACCGGCACCATCGGCAAGCGCTTCGCCAGCATCAATGTTGAGAA 198
   |||  :::::  |||  :::::  |||  :::::  |||  :::::  |||
89 SerArg.....LeuGlyAsnTyr 94
   |||  :::::  |||  :::::  |||
199 GTCGAGGACAAACCGCGTCGCCCTCGGTGAGCTCCTCTTCTGACCCCTG 248
   |||  :::::  |||  :::::  |||
94 rLeuProMetSerGlnLeu.....LeuT 103
   |||  :::::  |||  :::::  |||
249 GTGCCCTCCAGTACTCTCAGCGGTGTGATCCTCTTCGAGGACACCTCTAC 298
   |||  :::::  |||  :::::  |||
103 hrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyr.....Gln 117
   :::::  |||  :::::  |||  :::::  |||  :::::  |||
299 CAGAGCACCAAGGTTGGCAAGCCCTTCCTGACATCTCTCAAGAGGGCAA 348
   :::::  |||  :::::  |||  :::::  |||  :::::  |||
118 LeuProGluHisLeuTrpGlyGlnIleVal..... 128

```

```

|||||:|||||
349 CGTCTCCCGCATCAAGTGGACAAAGGTACCATCGAGCTCGCTGGAA 398
129 .....ProThrLeuHisLeuT 134
399 CCAATGGTGAGACACACCCAGCGCTTTGATGACCTTGGCAAGCGCTGC 448
134 yrglnaspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
449 GCCAAGTACTA.....CGAGGCTGGTGGCCG.....CTCGC 480
151 SerValTyrArgAsnProGlu..... 157
481 CAAGTGGCGTGGCTGCTCAAGATCGGTGCCACGAGCCATCGCAACTCT 530
158 ....LeuAsnGlnCysAlaGlyAlaAlaMetSerLysHisLeuThra 173
531 CCATCGACCAAGATGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580
173 snSerAlaIleAspIleTyrValProAspLeuGluIleLysSerGlnAla 189
581 GAAGAAGC.....GTCGTGTGCCATCGTTGAGCCTGAATCC..... 618
190 LeuTyrGluLeuGlnAsnArgLeuCysGlnTyrTrpLeuGlu 203
619 .....TGGTGGATGGAC 630

```

seq_name: gb_gss:BH070893

```

seq_documentation_block:
LOCUS BH070893
DEFINITION RPCI-24-275C20-TV RPCI-24 Mus musculus genomic clone RPCI-24-275C20
, DNA sequence.
ACCESSION BH070893
VERSION BH070893.1 GI:14890490
KEYWORDS GSS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 533)
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tldb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 275 row: C column: 20
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1..533
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-275C20"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI; The
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the

```

FEATURES

source

BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT	146 a	144 c	127 g	116 t
ORIGIN				

alignment_scores:

Quality:	89.00	Length:	87
Ratio:	1.894	Gaps:	3
Percent Similarity:	54.023	Percent identity:	32.184

alignment_block:

US-09-674-779-2 x BH070893

Align seg 1/1 to: BH070893 from: 1 to: 533

```

23 ProfileProThrAsnPro.....GlnValSerProIleLysThrProSe 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
64 CCTATTCACACAGCCACCTCCAGATAGTCCACTCCCTGGTCCGAA 113
37 rValLeuIleThrLysAspLysIleGlyAspHisHisThrHisGluHisA 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
114 AATATACAAACCATCATCAGCATATATACACAGACATGCATGCCACATA 163
54 spGluSerValSerHisValGlyLeuGlnAlaHisPheGluThr..... 68
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 CAGAGGACATAGTACGCTGGGATCCTCAGTTCTTCTCTACTGTCTCC 213
69 .....Trp.LeuGlnMethHisAlaThrLysGlnGluValV 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 TTTCTCTCTTATGTCTTCTCTGCTGCTCATCCACCAAGCCAAAGCAG 263
81 alArgTyrGlnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProPro 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 TACGTGGCCCCCAGCAGTGCAGCACCGTCTGCTGTCTTCTAGAACTG 313
98 MetSerGln 100
|||||
314 TGAAGTCAG 322

```

seq_name: gb_estl:BB692619

seq_documentation_block:

```

LOCUS BB692619
DEFINITION BB692619 RIKEN full-length enriched, 2 days neonate sympathetic
ganglion Mus musculus cDNA clone 7120437K23 3', mRNA sequence.
ACCESSION BB692619
VERSION BB692619.1 GI:16019352
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 550)
Hayatsu,N., Hiramoto,K., Hiraoka,T., Furuno,M., Hanagaki,T.,
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,
Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscl.riken.go.jp,

```

TITLE

JOURNAL

COMMENT


```

alignment_scores:
  Quality: 89.00      Length: 185
  Ratio: 1.035       Gaps: 12
  Percent Similarity: 46.486   Percent Identity: 27.027

alignment_block:
US-09-674-779-2 x BG207133/rev ..

Align seg 1/1 to reverse of: BG207133 from: 1 to: 647

25 ProThrAsnProGlnValSerProIleLysThrProSerVal.....Le 39
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
632 CCAGAGAAGCCAG.....ACCCCTGAAGTAGTGGAT 598

39 uileThrLysAspLysIleGlyAspHisThrHis....GluHisAspG 55
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
597 TGCCACATGACGGCTCCATGTGGACAGCCAGCAAGCACTTCTAGCTCCACAGTG 548

55 luSerValSerHisValGlyLeuGlnAlaHisPheGluThrTrpLeuGln 71
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
547 AGAGCCTATCAACATAGTACGTACAGTACCATGACCATGACCCCTGGCTTCAG 498

72 MethHisAlaThrLysGlnValValArgTyrGlnAlaTyrLeuG1 88
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
497 CTC.....CCAGTGGGAACAGCAGCCCAACAGCCTGGGAATACATCTCT 454

88 nSerArgLeuGlyAsnTyrLeuProPomMetSerGlnLeuThrThra 105
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
453 TTCCAGAGCCAGCGCTGCATCCAGTCCCATGCCATGCTATACCTACTG. 405

105 laArgSerTrpGlnAlaCys..... 111
   |||
404 .....TACTGGACACCCCTGCTTCTAGTCCACAGTGAGCCCTATCAAC 360

112 .....GlyHisGluProTyr...GlnLeuProProGluH1 122
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
359 TAGTAAGTACGAGTACGATGAGCCCTGCTTCCAGTCCAGTGGAGCA 310

122 sLeuTrpGlyGlnIleValProThrLeuHisLeuTyrGlnAspLeu...L 138
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
309 G.....CAGCAACACCCCTGGGAATACATCTCTCTTCCA 275

138 ysSerArgGlyIleLeuProAlaAsnThrGlnIleArgSerValTyr... 153
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
274 AGAGCCAGCTGCTACCATCCAGTCCCATGCCATGCTATACCTACTGTCTG 225

154 .....ArgAsnProGluLeuAsnGln.....CysAl 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
224 ACACCCCTGCTGTGAAGAAATCCAGAGTGACCAAGACCAAGTCTTGCTC 175

162 agLYGlyAlaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleT 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
174 TTATGGAACTTCTATT.....CTAGGAATGGGAATGGTGTATT 134

179 rpVal 180
   |||:
133 ATATA 129

```

```
seq_name: gb_gss:CNS02K1D
```

```
seq_documentation_block:
```

```

LOCUS      CNS02K1D              1023 bp      DNA      linear      GSS 14-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            144J11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL200938
VERSION    AL200938.1  GI:7859283
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1023)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1023)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1023)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
    source
        1..1023
            /organism="Tetraodon nigroviridis"
            /db_xref="taxon:99883"
            /clone="144J11"
            /clone_lib="G"
            /notes="Genoscope sequence ID : COAG144CE06SP1-end :
                PUC-Ori"
BASE COUNT      232 a      273 c      271 g      242 t      5 others
ORIGIN

```

```
alignment_scores:
```

```

  Quality: 89.00      Length: 306
  Ratio: 0.761       Gaps: 19
  Percent Similarity: 38.235   Percent Identity: 20.915

```

```
alignment_block:
```

```
US-09-674-779-2 x CNS02K1D ..
```

```
Align seg 1/1 to: CNS02K1D from: 1 to: 1023
```

```
23 ProfileProThrAsnPro.....GlnValSerProIleLysThrPr 36
   ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
225 CCCCCGCCATCAACCCCTGCTCTTAAGCAGGATTGGCCCAT..... 266

```

```
36 oSerValLeulleThrLysAspLysIleGlyAspHis..... 48
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
267 .....ATTGGAGACCACCCCTGTAGGAGGAT 291

```

```
49 .....HisThrHisGluHisAspGluSer..... 56
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
292 GCACACACACACACACACACACACACACACCGCTTTGTGATAGTGTGTT 341

```

```
56 ..... 56
342 ACTGTATACGGCGGCGCGGCGGTTACGCTTGTAGTYAGCCTCCAGCCCTG 391

```

```
57 .....ValSerHisValGlyLeuGlnAlaHisPheGlu..... 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
392 CGTAAAGCTCCTGTTCAGTATCGACTCTACCGGGTTATTTTCAGCCAC 441

```

```
68 .....Thr 68
   |||
442 CGAATAATGAACACCTGGATCAGCTAATAGTTGGGGGTTCGGAGCTACT 491

```

```
69 TrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrGlnAl 85
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```


9-674-779-2 x BF177880

[illegible]

Align seg 1/1 to: BF177880 from: 1 to: 509

```

49 HisThrHisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHi 65
   ||| |||:||||| :||| ||| ||| ||| ||| ||| ||| ||| |||
97 CACCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 146
   ||| |||:||||| :||| ||| ||| ||| ||| ||| ||| ||| |||
65 sPheGluThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValAla 82
   ||| |||:||||| :||| ||| ||| ||| ||| ||| ||| ||| |||
147 CTGCCCCCACCCTCCAGCACTCCACCAC..... 174
82 rgTyrglnAlaTyrglnGlnSerArgLeuGlyAsnTyrglnProMet 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 ..CAGCAGCTACCGCTCTCCACTACTCCAGCAGCTGCACCCCTG 222
99 SerGlnLeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluPr 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
223 ..... 236
115 oTyrglnLeuProGluHisLeuTrpGlyGlnIleValProThrLeuH 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
237 CTTCCAGCTCCCACTCAACCGCTTCCCCACCGGTGCGACCCCTTC 286
132 isLeuTyrglnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGln 148
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 CGCTT.....CTCTCCACCGCGCCCAATCCATCCCGCCAGCATCAG 330
149 IleArgSerValTyrglnAsnProGluLeuAsnGlnCysAlaGlyClyAl 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
331 GCCCAGGCC.....CGCGCGTGGACACG....GCCCAACACG 365
165 aaAlaMetSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProA 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
366 GCTGATACACCACTC..... 388
182 spLeuGlu..IleLysSerGlnAlaLeuTyrglnLeuGlnAsnArgLeuCy 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 CGCTGCAATTCTC...ACTCAACAGCAGCCATCATTCGTGCGCAGCTCTTG 435
198 sGlnTyrglnLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuT 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 TTGCTACCTCTCTGCGCATGGT.....CTGCGCTTAGGATCTCT 473
215 yrAlaThrGlyAlaIle 220
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474 TCCAGGTGGAGGCTTG 490

```

seq_name: gb_est2:BM002719

seq_documentation_block:

LOCUS BM002719 523 bp mRNA linear EST 25-OCT-2001
 DEFINITION 1031106C05.x1 C. reinhardtii CC-1690, Stress II (normalized),
 Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BM002719

VERSION BM002719.1 GI:16437499

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 523)

Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre

, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1031

Unpublished (2001)

Contact: Charles Hauser

DCMB Box 91000

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

Location/Qualifiers

FEATURES

source

```

1..523
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized)
), Lambda Zap II"
/notes="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda zap clones by superinfection with Exassist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."

```

BASE COUNT 156 a 163 c 112.9 92 t

ORIGIN

alignment_scores:

Quality: 88.50 Length: 84

Ratio: 1.735 Gaps: 5

Percent similarity: 60.714 Percent identity: 33.333

alignment_block:

US-09-674-779-2 x BM002719 ..

Align seg 1/1 to: BM002719 from: 1 to: 523

```

60 ValGlyLeuGlnAlaHisPheGluThrTrpLeuGlnMethHisAlaTh 76
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 ATTGCCCAACAAGCACATTCGACACA.....CAGCAAAACGACGAGC 226
76 rLysGlnGluValValArgTyrglnAlaTyrglnSerArgLeuGlyA 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 AGAACAAACACGCTATCCGCCCAAGACACA.....TCAATCGCGCTGGCA 270
93 snTyrglnProMetSerGlnLeuLeuThr...ThrAlaArgSerTrp 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
271 CC.....ATCAATGACCAAACTACAAAAGACGCGGAGCTGG 314
109 GlnAlaCysGly.....HisGluProTyrglnLeuProGluHisLe 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315 CGAAACTGCGCTCAAGTTCACAGCGCAGCGTGCACATGTGCG 364
123 utrGlyGlnIleValProThrLeuHisLeuTyrglnAspLeuLysSerA 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 CTGCTGCACGCCCTATCCGAAACAGCAGCTACCGTGTGTCGCCGACA 414
140 rg 140
   ||
415 GA 416

```

seq_name: gb_est1:AW757658

seq_documentation_block:

LOCUS AW757658 546 bp mRNA linear EST 03-MAY-2000
 DEFINITION 874002G08.x1 C. reinhardtii CC-1690, Lambda zap II Chlamydomonas
 reinhardtii cDNA, mRNA sequence.

ACCESSION AW757658

VERSION AW757658.1 GI:7687010

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii


```
138 sSerArgGly...IleLeuProIaAsnThrGlnIleArgSerValTyrA 154
|:::||||| |:::||||| |:::|||||
312 GACGAGAGGTCAGAGACTTCCTGCTATTACCAGTC..... 277
154 rGAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMetSerLysHis 170
|:::||||| |:::||||| |:::|||||
276 .....CTCACTCATAGGCGAGGGACCTGGCTCGAGAGAAAGCAT 238
171 LeuThrAsnSerAlaIleAspIleTyrValProAspLeuGluIleLysSe 187
|:::||||| |:::||||| |:::|||||
237 GCTGAGGATGCTAGGATCCACTC.....ATCCTCACTCACAGAGT 197
187 rGlnAlaLeuTyrGlu...LeuGlnAsnArgLeuCysGlnTyrTrpLeuG 203
|:::||||| |:::||||| |:::|||||
196 AGAAGCTCTACACAGAAAGGCAAGACAGAGGAAGTGCAAA..... 157
203 luHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaThrGlyAla 219
|:::||||| |:::||||| |:::|||||
156 .....GCTGCTACCCCTGTC 142
220 IleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnPheSerGI 236
|:::||||| |:::||||| |:::|||||
141 TGCACCCCTGAACATAAA.....GCAGAGTGTCACT 110
236 uThrAsnSerIleCysArgHisValLeuProLysAsnLys 249
|:::||||| |:::||||| |:::|||||
109 TAGTGAGAGCTGCACCACCACCATCTCACCCTCCAGCTCAAGA 70
```

OM of: US-09-674-779-2 to: N_Geneseq_032802.* out_format : pfs
 Date: Sep 16, 2002 8:00 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frameat_p2n.model -DEV=xlh
 -Q/cgnt2_1/USPTO.spool/US09674779/runat_12092002_124144_26926/app_query.fasta_1.310
 -DB=N_Geneseq_032802 -QPM=fastap -SUFFIX=oli.rng -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -DELOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=quality
 -THRN_MIN=1 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
 -USPR=US09674779_@CGN1_1_165 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
 Search information block:
 Query: US-09-674-779-2
 Query length: 250
 Database: N_Geneseq_032802.*
 Database sequences: 1736436
 Database length: 858457221
 Search time (sec): 199.450000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
 WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strid Orig	ZScore	Escore	Len	Documentation
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59797 +	..	250.00	4727.15	5.9e-2	
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF28530 +	250.00	4701.71	1.5e-2		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF193830 +	9.00	153.44	3.37		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59798 +	8.00	152.79	3.66		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59799 +	8.00	149.78	5.39		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA07054 +	8.00	135.32	34.41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK88795 +	8.00	135.32	34.41		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAQ10230 +	8.00	123.74	97.80		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK08548 +	8.00	123.48	152.07		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH17512 +	8.00	123.48	157.11		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAQ10484 +	8.00	122.46	179.20		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK46451 +	8.00	115.26	450.98		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:ABLI18726 +	8.00	102.35	2.4e+0		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ01425 +	8.00	77.38	5.8e+04		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI31302 +	7.00	131.99	52.76		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA67992 +	7.00	128.11	86.74		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI22900 +	7.00	128.11	86.74		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA18813 +	7.00	124.55	136.97		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAO05959 +	7.00	123.24	162.17		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT:AAK51506 +	7.00	123.02	164.96		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAK14154 +	7.00	121.56	200.99		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA73240 +	7.00	120.95	217.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK21676 +	7.00	120.95	217.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK47841 +	7.00	120.95	217.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI53670 +	7.00	120.95	217.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI12381 +	7.00	119.85	250.35		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI21257 +	7.00	119.85	250.35		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK65002 +	7.00	119.79	252.17		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAK89764 +	7.00	119.74	253.99		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA50979 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA68959 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA35907 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK17286 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK43076 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI23842 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI49154 +	7.00	119.30	268.51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI09449 +	7.00	119.30	268.51		

/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT:AAK61584 + 7.00 119.23 271.
 /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA11586 - 7.00 118.97 280
 /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV37035 + 7.00 118.55 295.
 /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH01713 + 7.00 118.55 295.
 /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:AAI16270 + 7.00 118.41 301.
 /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH01012 + 7.00 118.09 313
 /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH01298 + 7.00 118.00 317

seq name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59797

seq_documentation_block:

ID AAF59797 standard; DNA; 753 BP.

XX AAF59797;

XX 04-MAY-2001 (first entry)

XX Moraxella catarrhalis strain ATCC43617 BASB120 DNA.

KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;

KW genetic immunisation; infection; upper respiratory tract; otitis media;

KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;

KW invasive disease; antibacterial; auditory; ds.

OS Moraxella catarrhalis.

XX WO200109335-A2.

XX 08-FEB-2001.

XX 31-JUL-2000; 2000WO-EP07361.

XX 03-AUG-1999; 99GB-0018281.

XX (SMTK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonnard J;

XX WPI: 2001-159872/16.

XX P-PSDB; AAB60645.

New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis strain American Type Culture Collection 43617, for use as therapeutic agents or vaccines against bacterial infections, e.g. otitis media or pneumonia

Claim 13; Page 64; 75pp; English.

The invention relates to the Moraxella catarrhalis strain ATCC43617 BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The invention also relates to immunogenic fragments of the BASB120 protein, expression vectors and host cells comprising BASB120 nucleic acids, the recombinant production of BASB120, vaccine compositions comprising the BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic compositions comprising the anti-BASB120 antibody, and a method of identifying a Moraxella catarrhalis infection via the detection of BASB120 proteins or antibodies. The vaccine compositions of the invention are useful as prophylactic or therapeutic agents against Moraxella catarrhalis infections in mammals, particularly humans. Moraxella catarrhalis is a Gram negative bacterium frequently isolated from the human upper respiratory tract, which is responsible for several pathological conditions. It is responsible for about 15% of otitis media cases in children (which can lead to temporary or permanent hearing loss). It also causes pneumonia in elderly people, and sinusitis, nosocomial infections and, less frequently, invasive diseases. BASB120 proteins or nucleotides may additionally be used in screening for novel antibacterial compounds, and in the diagnosis and staging of infections. The present sequence represents DNA encoding the Moraxella catarrhalis strain ATCC43617 BASB120 protein.

Sequence 753 BP; 239 A; 166 C; 152 G; 196 T; 0 other;

```

alignment_scores:
  Quality: 250.00      Length: 250
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAF59797  ..

Align seg 1/1 to: AAF59797 from: 1 to: 753

1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
1 ATGAAAAATTTTAACTCACTACTTATACTACACTTATCAGCAGTATGCT 50

17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
1 GGTTCATGCAGTCACCCATACCAATCCACAGATATCCCAATAA 100

34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisThr 50
101 AAACGCCATCGTACTGATTACTAAAGATAAAATCGGTGATCATATACA 150

51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
151 CATGACGACGATGAATCTGTAAAGCCATGTCGGTTGCAGGCACATTTGA 200

67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
201 GACTTGGCTACAGATGCACCATGCCACCAACAGAGGTAGTAGGATC 250

84 InAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
251 AGGCGTATTTACAATCAAGACTTGGTAATTATCTGCCACCAATGAGTCAA 300

101 LeuLeuThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG1 117
301 CTACTAACGACTGCACGTCATGCAGGCATGTGGTCATGAACCTTATCA 350

117 nLeuProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeuT 134
351 GTCGCCACGACAGACATCTTTGGGTGCAGATGTACCAACATGTCACITGT 400

134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
401 ATCAAGATCTCAAAAGTAGGGGCATATTGCCAGCAAAATACCCAAATTCGC 450

151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
451 TCAGTTTATCGCAATCCTGAACTCAACCAATGTGCTGGTGTGCAGCTAT 500

167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
501 GAGTAAGCATTTGACCAATAGTGCATTTGATATTTGGGTGCTGACCTTG 550

184 luIleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuCysGlnTyr 200
551 AAATAAAAAAGCCAGGCACCTGTAGCTTCAAAAACCGCCTATGCCAATAT 600

201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaTh 217
601 TGGCTAGACATGCGGAACCAAAATTTTGGGCTGGGTTTATACGGCAC 650

217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
651 AGGGCGGATTCATCTGGATACCCAGGGTTTAGAAAATGGGGTCTCAAT 700

234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
701 TTCTCTGAAACAACTCTATTGTGCTGCTATGCTTACCAAAAAATAAGTGA 750

seq_name: /SID1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF28530
seq_documentation_block:

```

```

ID AAF28530 standard; DNA; 23210 BP.
XX
AC AAF28530;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #17.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
XX
DR WPI; 2001-041427/05.
XX
XX Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX
PS Claim 1; Page 141-146; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 23210 BP; 7108 A; 4482 C; 4951 G; 6668 T; 1 other;

alignment_scores:
  Quality: 250.00      Length: 250
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAF28530  ..

Align seg 1/1 to: AAF28530 from: 1 to: 23210

1 MetLysAsnPheAsnGlnTyrPheIleThrThrLeuIleSerSerMetLe 17
11483 ATGAAAAATTTTAACTCACTACTTATACTACACTTATCAGCAGTATGCT 11532

17 uValAlaCysSerAlaProIleProThrAsnProGlnValSerProIleL 34
11533 GGTTCATGCAGTCACCCATACCAATCCACAGATATCCCAATAA 11582

34 ysThrProSerValLeuIleThrLysAspLysIleGlyAspHisThr 50
11583 AAACGCCATCGGTACTGATTACTAAAGATAAAATCGGTGATCATATACA 11632

51 HisGluHisAspGluSerValSerHisValGlyLeuGlnAlaHisPheG1 67
11633 CATGACGACGATGAATCTGTAAAGCCATGTCGGTTGCAGGCACATTTTGA 11682

```

```

67 uThrTrpLeuGlnMetHisHisAlaThrLysGlnGluValValArgTyrG 84
11683 GACTTGGCTACAGATGCACCATGCCACCAACAAGAGGTAGTTAGCTATC 11732
84 lnAlaTyrLeuGlnSerArgLeuGlyAsnTyrLeuProProMetSerGln 100
11733 AGCGTATTACAAATCAAGACTTGGTAATATCTGCGCAACATGAGTCAA 11782
101 LeuLeuThrThrAlaArgSerTrpGlnAlaCysGlyHisGluProTyrG 117
11783 CTACTACGACTGCACGGTCATGCGAGCGATGGTCATGAACCTTATCA 11832
117 nLeuProProGluHisLeuTrpGlyGlnIleValProThrLeuHisLeu 134
11833 GCTGCCACCAAGACATCTTTGGGTGAGATTTGACCAACATGCACTTGT 11882
134 yrGlnAspLeuLysSerArgGlyIleLeuProAlaAsnThrGlnIleArg 150
11883 ATCAAGATCTCAAAAGTAGGGGCATATGCGACAAATACCCAAATCGC 11932
151 SerValTyrArgAsnProGluLeuAsnGlnCysAlaGlyGlyAlaAlaMe 167
11933 TCAGTTATGCGCAATCCTGAACCTACCAATGCTGGTGGTGCACCTAT 11982
167 tSerLysHisLeuThrAsnSerAlaIleAspIleTrpValProAspLeuG 184
11983 GAGTAAGCATTTGACCAATAGTGCATATGATATTTGGGTGCGCTGACCTG 12032
184 lulleLysSerGlnAlaLeuTyrGluLeuGlnAsnArgLeuGlnTyr 200
12033 AATAAAGCCAGGACATGATGAGCTTCAAAACCGCTATGCCAATAT 12082
201 TrpLeuGluHisGlyGluAsnGlnAsnPheGlyLeuGlyLeuTyrAlaIth 217
12083 TGGCTAGAGCATGGCGAAACCAAAATTTGGGCTGGGTTTATACGCCAC 12132
217 rGlyAlaIleHisLeuAspThrGlnGlyPheArgLysTrpGlyAlaGlnP 234
12133 AGGGCGGATTCATCTGCATACCAAGGGTTAGAAAATGGGGTGTCTCAAT 12182
234 heSerGluThrAsnSerIleCysArgHisValLeuProLysAsnLysLeu 250
12183 TTCTGTAACAACTCTATTGTCGTCATGCTTACCAAAAAATAGCTA 12232
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI93830
seq_documentation_block:
ID AAI93830 standard; cDNA; 474 BP.
XX
AC AAI93830;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 13890.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.

```

Tang YT, Liu C, Drmanac RT;
WPI; 2001-514838/56.
P-PSDB; AAO13899.
Isolated nucleic acids and polypeptides, useful for preventing
diagnosing and treating e.g. leukaemia, inflammation and immune
disorders -
Claim 1; SEQ ID NO 13890; 1399pp + Sequence Listing; English.
The invention relates to human polynucleotides (AAI79941-AAI93841) and
the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
cytokine cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 474 BP; 116 A; 109 C; 139 G; 110 T; 0 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAI93830 ..
Align seg 1/1 to: AAI93830 from: 1 to: 474

33 lleylsthPrSerValLeuIleThr 41
|||||
359 ATCAAGACCCCATCTGTACTATAACG 385

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59798
seq_documentation_block:
ID AAF59798 standard; DNA; 40 BP.
XX
AC AAF59798;
XX
DT 04-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis BASB120 mutagenic PCR primer, SEQ ID NO:3.
XX
KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; mutagenic PCR primer; ss.
XX
OS Moraxella catarrhalis.
OS Synthetic.
XX
PN WO200109335-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07361.
XX
PR 03-AUG-1999; 99GB-0018281.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Thonnard J;
 XX WPI; 2001-159872/16.
 XX
 XX New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
 PT strain American Type Culture Collection 43617, for use as therapeutic
 PT agents or vaccines against bacterial infections, e.g. otitis media or
 PT pneumonia
 XX
 XX Example 2; Page 51; 75pp; English.
 PS
 XX The invention relates to the Moraxella catarrhalis strain ATCC43617
 CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
 CC invention also relates to immunogenic fragments of the BASB120 protein,
 CC expression vectors and host cells comprising BASB120 nucleic acids, the
 CC recombinant production of BASB120, vaccine compositions comprising the
 CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
 CC compositions comprising the anti-BASB120 antibody, and a method of
 CC identifying a Moraxella catarrhalis infection via the detection of
 CC BASB120 proteins or antibodies. The vaccine compositions of the invention
 CC are useful as prophylactic or therapeutic agents against Moraxella
 CC catarrhalis infections in mammals, particularly humans. Moraxella
 CC catarrhalis is a Gram negative bacterium frequently isolated from the
 CC human upper respiratory tract, which is responsible for several
 CC pathological conditions. It is responsible for about 15% of otitis media
 CC cases in children (which can lead to temporary or permanent hearing
 CC loss). It also causes pneumonia in elderly people, and sinusitis.
 CC nosocomial infections and, less frequently, invasive diseases. BASB120
 CC proteins or nucleotides may additionally be used in screening for novel
 CC antibacterial compounds, and in the diagnosis and staging of infections.
 CC Sequences AAF59798-AAF59799 represent PCR primers used to introduce
 CC EcoRI and SalI restriction sites into Moraxella catarrhalis strain
 CC ATCC43617 BASB120 DNA to facilitate cloning into the pTL22 vector.
 XX
 XX Sequence 40 BP; 17 A; 4 C; 7 G; 12 T; 0 other;
 SQ

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF59798 ..
 Align seg 1/1 to: AAF59798 from: 1 to: 40
 1 MetLysAsnPheAsnGlnTyrPhe 8
 |||||||||||||||||||||||||
 16 ATGAAAAATTTTAAATCAATACACTT 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAF59799

seq_documentation_block:
 ID AAF59799 standard; DNA; 60 BP.
 XX
 AC AAF59799;
 XX
 DT 04-MAY-2001 (first entry)
 XX
 DE Moraxella catarrhalis BASB120 mutagenic PCR primer, SEQ ID NO:4.
 XX
 KW BASB120 protein; strain ATCC43617; antigen; antibody; vaccine;
 KW genetic immunisation; infection; upper respiratory tract; otitis media;
 KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
 KW invasive disease; antibacterial; auditory; mutagenic PCR primer; ss.
 XX
 OS Moraxella catarrhalis.
 OS Synthetic.
 XX
 PN WO200109335-A2.
 XX
 PD 08-FEB-2001.

XX 31-JUL-2000; 2000WO-EP07361.
 XX
 XX 03-AUG-1999; 95GB-0018281.
 XX (SMKT) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Thonnard J;
 XX
 XX WPI; 2001-159872/16.
 XX
 XX New BASB120 polypeptides and polynucleotides from Moraxella catarrhalis
 PT strain American Type Culture Collection 43617, for use as therapeutic
 PT agents or vaccines against bacterial infections, e.g. otitis media or
 PT pneumonia
 XX
 XX Example 2; Page 51; 75pp; English.
 PS
 XX The invention relates to the Moraxella catarrhalis strain ATCC43617
 CC BASB120 protein (AAB60645) and to DNA encoding it (AAF59797). The
 CC invention also relates to immunogenic fragments of the BASB120 protein,
 CC expression vectors and host cells comprising BASB120 nucleic acids, the
 CC recombinant production of BASB120, vaccine compositions comprising the
 CC BASB120 protein or nucleic acid, an antibody against BASB120, therapeutic
 CC compositions comprising the anti-BASB120 antibody, and a method of
 CC identifying a Moraxella catarrhalis infection via the detection of
 CC BASB120 proteins or antibodies. The vaccine compositions of the invention
 CC are useful as prophylactic or therapeutic agents against Moraxella
 CC catarrhalis infections in mammals, particularly humans. Moraxella
 CC catarrhalis is a Gram negative bacterium frequently isolated from the
 CC human upper respiratory tract, which is responsible for about 15% of otitis media
 CC pathological conditions. It is responsible for about 15% of otitis media
 CC cases in children (which can lead to temporary or permanent hearing
 CC loss). It also causes pneumonia in elderly people, and sinusitis.
 CC nosocomial infections and, less frequently, invasive diseases. BASB120
 CC proteins or nucleotides may additionally be used in screening for novel
 CC antibacterial compounds, and in the diagnosis and staging of infections.
 CC Sequences AAF59798-AAF59799 represent PCR primers used to introduce
 CC EcoRI and SalI restriction sites into Moraxella catarrhalis strain
 CC ATCC43617 BASB120 DNA to facilitate cloning into the pTL22 vector.
 XX
 XX Sequence 60 BP; 14 A; 5 C; 21 G; 20 T; 0 other;
 SQ

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAF59799/rev ..
 Align seg 1/1 to reverse of: AAF59799 from: 1 to: 60
 243 HisValLeuProLysAsnLysLeu 250
 |||||||||||||||||||||||||
 60 CATGCTCTTACCAAAATAAGCTA 37

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA07054

seq_documentation_block:
 ID ABA07054 standard; cDNA; 421 BP.
 XX
 AC ABA07054;
 XX
 DT 14-JAN-2002 (first entry)
 XX
 DE Human pancreatic cancer related cDNA, SEQ ID NO: 183.
 XX
 KW Human; cytostatic; antidiabetic; antiinflammatory; gastric; osteopathic;
 KW antihormone; antiulcer; thyroid-active; gene therapy; antisense therapy;
 KW pancreatic cancer antigen inhibitor; pancreatic cancer; pancreatitis;
 KW diabetes; endocrine disorder; acromegaly; hyperthyroidism;

KW gastrointestinal disorder; Crohn's disease; duodenal ulcer; ss.

XX Homo sapiens.

OS WO200155206-A1.

PN XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01353.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226581.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251031.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.

```

XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457717/49.
DR P-PSDB; ABB10724.
XX
XX Isolated pancreatic cancer polypeptide for treating, preventing and/or
PT prognosing disorders related to the pancreas including pancreatic
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID NO 183; 537pp; English.
XX
XX The invention relates to an isolated polypeptide comprising an amino
CC acid sequence at least 90% identical to 188 amino acid sequences fully
CC defined in the specification and encoded by 188 cDNA clones fully
CC defined in the specification. The invention also relates to a fragment
CC having biological activity, a domain, an epitope, full length protein,
CC variant, allelic variant or a species homologue of the fully defined
CC sequence. The polynucleotide and polypeptide are useful for treating,
CC preventing and/or prognosing disorders related to the pancreas including
CC pancreatic cancer, pancreatitis, diabetes, endocrine disorders such as
CC acromegaly or hyperthyroidism, and gastrointestinal disorders such as
CC Crohn's disease and duodenal ulcers. The present sequence is a
CC pancreatic cancer-related polynucleotide of the invention.
XX
SQ Sequence 421 BP; 124 A; 79 C; 91 G; 126 T; 1 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x ABA07054 ..
Align seg 1/1 to: ABA07054 from: 1 to: 421

211 GlyLeuGlyLeuTyrAlaThrGly 218
|||||
384 GGTGTGGGCTTATGCCACTGG 407

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-embl/NA2001A.DAT:AAK88795
seq_documentation_block:
ID AAK88795 standard; cDNA; 421 BP.
XX
XX AAK88795;
AC
XX
XX 05-NOV-2001 (first entry)
DT
XX
XX Human digestive system antigen coding sequence SEQ ID NO: 1111.
DE
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200155314-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01324.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR
XX
XX 04-FEB-2000; 2000US-0180628.
PR
XX
XX 24-FEB-2000; 2000US-0184664.
PR
XX
XX 02-MAR-2000; 2000US-0186350.
PR
XX
XX 16-MAR-2000; 2000US-0189874.
PR
XX
XX 17-MAR-2000; 2000US-0190076.
PR

```

```

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

```


CC pKT2-2 and used to transform E.coli JM109 which was cultured to
 CC produce streptokinase. See also AAR10195-R10200.
 XX
 SQ Sequence 1262 BP; 365 A; 350 C; 262 G; 285 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAQ10230/rev ..
 Align seg 1/1 to reverse of: AAQ10230 from: 1 to: 1262

192 GluLeuGlnAsnArgLeuCysGln 199
 |||||
 393 GAGTTACAGAACCGCTCTTTGTCAG 370

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAQ08548

seq_documentation_block:
 ID AAD08548 standard; DNA; 2007 BP.

XX
 AC AAD08548;

XX
 DT 04-SEP-2001 (first entry)

XX
 DE Feline calicivirus (FCV) capsid gene.

XX
 KW Raccoon poxvirus; RCNV; thymidine kinase; TK; vaccine; haemagglutinin;
 KW HA; feline pathogen; feline panleukopenia virus; FPV; feline calicivirus;
 KW FCV; capsid protein; virucide; ds.

XX
 OS Feline calicivirus.

XX
 FH Key Location/Qualifiers
 FT CDS 1..2007
 FT /*tag= a
 FT /product= "Feline calicivirus (FCV) capsid protein"

XX
 PN US6241989-B1.

XX
 PD 05-JUN-2001.

XX
 PF 03-NOV-1995; 95US-0552369.

XX
 PR 09-JUL-1991; 91US-0726609.

XX
 PR 27-JAN-1994; 94US-0190789.

XX
 PA (CORR) CORNELL RES FOUND INC.

XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX
 PI Scott FW, Ngichabe CK, Hu L, Esposito JJ;

XX
 DR WPI; 2001-407214/43.

XX
 DR P-PSDB; AAE04304.

XX
 PT New multivalent recombinant raccoon poxviruses, useful as vaccines to
 PT immunize felines against subsequent challenge by feline pathogens -

XX
 PS Example 6; Column 31-36; 35pp; English.

XX
 CC The present invention relates to multivalent recombinant raccoon
 CC poxviruses (RCNV), containing more than one exogenous gene inserted
 CC into either the thymidine kinase (TK) gene, the haemagglutinin (HA)
 CC gene, or a combination thereof. The multivalent recombinant raccoon
 CC poxviruses are useful as vaccines to immunise felines against subsequent
 CC challenge by feline pathogens. The recombinant multivalent vaccine is
 CC formed by inserting multiple genes such as a feline panleukopenia virus
 CC (FPV) gene, and/or a feline calicivirus (FCV) capsid protein gene each
 CC operably linked to a promoter, into a raccoon poxvirus for expression.

CC The present sequence is Feline calicivirus (FCV) capsid protein encoding
 CC gene.
 XX
 SQ Sequence 2007 BP; 528 A; 473 C; 421 G; 585 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAD08548/rev ..
 Align seg 1/1 to reverse of: AAD08548 from: 1 to: 2007

192 GluLeuGlnAsnArgLeuCysGln 199
 |||||
 1090 GAATTACAAATCGGTATGTCAG 1067

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH17512

seq_documentation_block:
 ID AAH17512 standard; cDNA; 2077 BP.

XX
 AC AAH17512;

XX
 DT 26-JUN-2001 (first entry)

XX
 DE Human cDNA sequence SEQ ID NO:16985.

XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX
 OS Homo sapiens.

XX
 PN EP1074617-A2.

XX
 PD 07-FEB-2001.

XX
 PF 28-JUL-2000; 2000EP-0116126.

XX
 PR 29-JUL-1999; 99JP-0248036.

XX
 PR 27-AUG-1999; 95JP-0300253.

XX
 PR 11-JAN-2000; 2000JP-0118776.

XX
 PR 02-MAY-2000; 2000JP-0183767.

XX
 PR 09-JUN-2000; 2000JP-0241899.

XX
 PA (HELI-) HELIX RES INST.

XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX
 DR WPI; 2001-318749/34.

XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

XX
 PS Claim 8; SEQ ID 16985; 2537pp + CD ROM; English.

XX
 CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 2077 BP; 630 A; 452 C; 395 G; 600 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAH17512/rev ..

Align seg 1/1 to reverse of: AAH17512 from: 1 to: 2077

211 GlyLeuGlyLeuTyAlaThrGly 218
 |||||
 937 GGACTAGGACTTTATGCCACTGGA 914

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:AAQ10484

seq_documentation_block:

ID AAQ10484 standard; cDNA; 2385 BP.

XX
 AC AAQ10484;

XX
 DT 22-APR-1991 (first entry)

XX
 DE Encodes Feline calicivirus capsid protein.

XX
 KW feline calicivirus capsid protein; FCV; multivalent vaccine; cat; ds.

XX
 OS feline calicivirus.

XX
 FH Key Location/Qualifiers

XX
 FT CDS 18..2024

XX
 FT FT /*tag= a

XX
 FT /product= capsid protein

XX
 PN W09101332-A.

XX
 PD 07-FEB-1991.

XX
 PF 09-JUL-1990; 90WO-US03753.

XX
 PR 21-JUL-1989; 89US-0383909.

XX
 PR 18-SEP-1989; 89US-0408989.

XX
 PA (UPJO) UPJOHN CO.

XX
 PI Wardley R, Post LE;

XX
 DR WPI; 1991-058117/08.

XX
 DR P-PSDB; AARI0686.

XX
 PT Feline calicivirus capsid protein - from recombinant DNA for subunit
 PT and multivalent vaccines for FCV

XX
 PS Claim 3; Page 28; 40pp; English.

XX
 CC CRPK cells were inoculated, total poly-A+ RNA was isolated and FCV
 CC double-stranded RNA was LiCl fractionated from this RNA. Double-
 CC stranded cDNA was prepared, blunt-ended with t4 DNA polymerase and
 CC treated with EcoRI methylase prior to ligation of EcoRI linkers. The

CC mixture was digested with EcoRI and ligated to lambda gt10. E. coli
 CC strains transformed with the viral vectors were plated and positive
 CC plaques were transferred to nitrocellulose. They were probed with a
 CC random-prime cDNA prepared from FCV genomic RNA. A plaque with a
 CC strong hybridisation signal was isolated and found to contain an
 CC EcoRI fragment (ca.4200bp). It was cloned in pUC18 to give pCV2.
 CC A second library was constructed in pUC18 using PstI linkers. A
 CC plasmid was selected from this library and designated pCV7. It was
 CC found to have restriction sites in common with pCV2. These two
 CC plasmids were used to isolate other regions of the FCV genome.

XX
 SQ Sequence 2385 BP; 642 A; 562 C; 502 G; 679 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAQ10484/rev ..

Align seg 1/1 to reverse of: AAQ10484 from: 1 to: 2385

192 GluLeuGlnAsnArgLeuCysGln 199

|||||

1107 GAATTACAAATCGTTATGTCAG 1084

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAAS46451

seq_documentation_block:

ID AAAS46451 standard; DNA; 6291 BP.

XX
 AC AAAS46451;

XX
 DT 18-DEC-2001 (first entry)

XX
 DE Tumour suppressor gene derived chemically modified sequence #173.

XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

XX
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.

XX
 OS Homo sapiens.

XX
 PN WO200168912-A2.

XX
 PD 20-SEP-2001.

XX
 PF 15-MAR-2001; 2001WO-EP02955.

XX
 PR 15-MAR-2000; 2000DE-1013847.

XX
 PR 06-APR-2000; 2000DE-1019058.

XX
 PR 07-APR-2000; 2000DE-1019173.

XX
 PR 30-JUN-2000; 2000DE-1032529.

XX
 PR 01-SEP-2000; 2000DE-1043826.

XX
 PA (EPIC-) EPIGENOMICS AG.

XX
 PI Olek A, Piepenbrock C, Berlin K;

XX
 DR WPI; 2001-602752/68.

XX
 PT Fragments of chemically modified genes associated with tumour suppressor

XX
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.

XX
 PT cancer

XX
 PS Claim 1; SEQ ID No 173; 27pp; English.

XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (cp DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and

CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6291 BP; 1501 A; 210 C; 1626 G; 2954 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAS46451/rev ..
 Align seg 1/1 to reverse of: AAS46451 from: 1 to: 6291
 34 LysThrProSerValLeuIleThr 41
 1884 AAACACCATCCGTTTAAATTACT 1861

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL18726

seq_documentation_block:
 ID ABL18726 standard; DNA; 35832 BP.

XX ABL18726;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7651.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions
 XX

PS Claim 1; SEQ ID NO 7651; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 35832 BP; 9547 A; 8198 C; 8102 G; 9985 T; 0 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABL18726 ..
 Align seg 1/1 to: ABL18726 from: 1 to: 35832
 8 PheIleThrThrLeuIleSerSer 15
 20967 TTCATTACTACGCTTATAAGCTCT 20990

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ01425

seq_documentation_block:
 ID AAZ01425 standard; DNA; 1038602 BP.

XX AAZ01425;
 XX
 DT 07-OCT-1999 (first entry)
 XX
 DE Complete genome sequence of Chlamydia trachomatis.
 XX
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; peritphatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WO9928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-371125/31.
 XX
 PT Genome sequence of Chlamydia trachomatis
 XX
 PS Claim 1; Page 373-656; 1755pp; English.
 XX

XX The present sequence represents the complete genome of Chlamydia
 CC trachomatis. Open reading frames (ORFs) of the genome encode
 CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 CC

CC be used to control growth of the microorganism. Chlamydia trachomatis is
 CC responsible for a large number of diseases, e.g. eye diseases such as
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
 CC conjunctivitis; genital diseases such as nongonococcal urethritis,
 CC epididymitis, cervicitis, salpingitis, perinephritis, Bartholinitis;
 CC pneumopathy in breast feeding infants; and venereal
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in
 CC treating these diseases.

SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAZ01425/rev ..
 Align seg 1/1 to reverse of: AAZ01425 from: 1 to: 1038602

212 LeuGlyLeuTyrAlaThrGlyAla 219
 |||||
 550259 CTAGGTTATACGCAACAGGAGCG 550236

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AA131302

seq_documentation_block:

ID AAL31302 standard; DNA; 51 BP.

AC AAL31302;

24-JAN-2002 (first entry)

Human SNP oligonucleotide #4510.

KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
 KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
 KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
 KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
 KW complement related protein; cytochrome; kinesin; cytokine; interferon;
 KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
 KW multifactorial disease; autoimmune disease; infection;
 KW nervous system disease; ss.

OS Homo sapiens.

PN WO200147944-A2.

PD 05-JUL-2001.

PF 28-DEC-2000; 2000WO-US35498.

PR 28-DEC-1999; 99US-0173419.

PR 27-DEC-2000; 2000US-0173419.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI; 2001-465210/50.

PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
 PT oncogenes and histones, useful for diagnosing and treating, e.g.
 PT cancer, autoimmune diseases and infections -

PS Claim 1; Page 2683; 4143pp; English.

CC The present invention relates to oligonucleotides encoding polymorphic
 CC variants of proteins related to amylases, amyloid proteins, angiotensin,
 CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
 CC histones, kinases, colony stimulating factors, complement related

CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
 CC G-protein coupled receptors and thioesterases. The present sequence is
 CC one such oligonucleotide. The oligonucleotides and the peptides encoded
 CC by them may be used in the prevention, diagnosis and treatment of
 CC diseases associated with inappropriate expression of the proteins listed
 CC above. Disorders that may be prevented, diagnosed and/or treated include
 CC multifactorial diseases with a genetic component, such as autoimmune
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
 CC systemic lupus erythematosus and Grave's disease), inflammation, cancer
 CC (e.g. cancers of the bladder, brain, breast, colon and kidney,
 CC leukaemia), diseases of the nervous system and an infection of pathogenic
 CC organisms.

SQ Sequence 51 BP; 10 A; 9 C; 16 G; 16 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAL31302/rev ..

Align seg 1/1 to reverse of: AAL31302 from: 1 to: 51

19 AlaCysSerAlaProIlePro 25

|||||

50 GCATGCTCAGCCCCCATCCCA 30

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:ABA67992

seq_documentation_block:

ID ABA67992 standard; DNA; 86 BP.

XX AC ABA67992;

XX 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #16297.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -

PS Claim 4; SEQ ID NO 16297; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 86 BP; 24 A; 12 C; 11 G; 39 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x ABA67992/rev ..

Align seg 1/1 to reverse of: ABA67992 from: 1 to: 86

98 MetSerGlnLeuThrThr 104

|||||
53 ATGTCACAACTTTTGACTACA 33

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI22900

seq_documentation_block:

ID AAI22900 standard; DNA; 86 BP.

XX AC AAI22900;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #12833 for gene expression analysis in human cervical cell sample.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX KW cervical cancer; ss.

XX OS Homo sapiens.

XX PN WO200157278-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00670.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DX, Chen W, Rank DR;

XX DR WPI; 2001-488901/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human cervical epithelial cells -

XX PS Claim 25; SEQ ID No 12833; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 86 BP; 24 A; 12 C; 11 G; 39 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAI22900/rev ..

Align seg 1/1 to reverse of: AAI22900 from: 1 to: 86

98 MetSerGlnLeuThrThr 104

|||||
53 ATGTCACAACTTTTGACTACA 33

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA18813

seq_documentation_block:

ID ABA18813 standard; DNA; 139 BP.

XX AC ABA18813;

XX DT 23-JAN-2002 (first entry)

XX DE Human nervous system related polynucleotide SEQ ID NO 11144.

XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskinking; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX OS Homo sapiens.

XX PN WO200159063-A2.

XX PD 16-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01334.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases.

Disclosure; SEQ ID NO 11144; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 139 BP; 31 A; 27 C; 24 G; 57 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x ABA18813/rev ..

Align seg 1/1 to reverse of: ABA18813 from: 1 to: 139

17 LeuValAlaCysSerAlaPro 23
|||||
109 TTAGTGGCGTGCAGTGCACCA 89

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT: AAC05959

seq_documentation_block:

ID AAC05959 standard; cDNA; 166 BP.

XX AC AAC05959;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 10034.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST) GENSET.

XX PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX DR WPI: 2000-500381/45.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX PS Claim 1; SEQ ID 10034; 71pp + CD-ROM; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.

XX SQ Sequence 166 BP; 41 A; 35 C; 36 G; 51 T; 3 other;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAC05959/rev ..
Align seg 1/1 to reverse of: AAC05959 from: 1 to: 166

141 GlyIleLeuProAlaAsnThr 147
|||||
122 GGCATACCTGCTGCAACACA 102

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT: AAX51506

seq_documentation_block:

ID AAX51506 standard; cDNA; 169 BP.

XX AC AAX51506;

XX DT 21-JUN-1999 (first entry)

XX DE Human secreted protein 5' EST SEQ ID NO:85.

XX KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
XX KW forensic; gene therapy; chromosome mapping; signal peptide;
XX KW upstream regulatory sequence; cytokine activity; cell proliferation;
XX KW differentiation; haematopoiesis regulation; tissue growth regulation;
XX KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
XX KW thrombolytic; anti-inflammatory; tumour inhibition; ds.

XX OS Homo sapiens.

XX PN WO9906549-A2.

XX PD 11-FEB-1999.

XX PF 31-JUL-1998; 98WO-IB01231.

XX PR 01-AUG-1997; 97US-0905279.

XX PA (GEST) GENSET.

XX PI Duclert A, Dumas Milne Edwards J, Lacroix B;

XX DR WPI: 1999-153779/13.

XX DR P-PSDB; AAY12728.

XX PT New nucleic acids encoding human secreted proteins - obtained from
XX PT cDNA libraries derived from testis, ovary, uterus and spleen tissue
XX PS Claim 1; Page 208; 522pp; English.

XX CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
XX CC human secreted proteins, and encode the proteins given in AAY12681 to
XX CC AAY12913, respectively. The proteins given represent the signal peptide
XX CC and an N-terminal fragment of a secreted protein. The nucleic acid
XX CC sequences can be used for producing secreted human gene products. They
XX CC can also be used to develop products for diagnosis and therapy. The
XX CC proteins obtained may have cytokine activity, cell
XX CC proliferation/differentiation activity, haematopoiesis regulating
XX CC activity, tissue growth regulating activity, reproductive hormone
XX CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and
XX CC thrombolytic activity, receptor/ ligand activity, anti-inflammatory
XX CC activity, tumour inhibition activity or other activities. The products
XX CC can be used in forensic, gene therapy and chromosome mapping procedures.
XX CC The sequences can also be used for obtaining corresponding promoter
XX CC sequences. The nucleic acids encoding the signal peptide can be used for
XX CC directing extracellular secretion of a polypeptide or the insertion of a
XX CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX SQ Sequence 169 BP; 42 A; 35 C; 38 G; 51 T; 3 other;

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAX51506/rev ..

Align seg 1/1 to reverse of: AAX51506 from: 1 to: 169

141 GlyIleLeuProAlaAsnThr 147
 |||||
 123 GGCATACTGCTGCACACACA 103

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAS65817

seq_documentation_block:

ID AAS65817 standard; cDNA; 171 BP.

AC AAS65817;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #1621.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

P-PSDB; ABG01630.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID No 1621; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 171 BP; 46 A; 39 C; 59 G; 27 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAS65817 ..

Align seg 1/1 to: AAS65817 from: 1 to: 171

162 AlaGlyGlyAlaAlaMetSer 168

|||||

115 GCAGGAGGGGCGCCATGAGT 135

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC14154

seq_documentation_block:

ID AAC14154 standard; cDNA; 208 BP.

XX AAC14154;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 18229.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

PN EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 18229; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dr primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

XX Sequence 208 BP; 62 A; 35 C; 56 G; 54 T; 1 other;

alignment_scores:

Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:
US-09-674-779-2 x AAC14154/rev ..
Align seg 1/1 to reverse of: AAC14154 from: 1 to: 208

30 ValSerProIleLysThrPro 36
|||||
126 GTGAGCCCAATTAAACTCCT 106

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA73240
seq_documentation_block:
ID ABA73240 standard; DNA; 226 BP.
XX
AC ABA73240;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #21545.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 21545; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x ABA73240 ..
Align seg 1/1 to: ABA73240 from: 1 to: 226

```

```

35 ThrProSerValIleThr 41
|||||
171 ACACCATCAGTATTATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK21676
seq_documentation_block:
ID AAK21676 standard; DNA; 226 BP.
XX
AC AAK21676;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 21667.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 21667; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAK21676 ..
Align seg 1/1 to: AAK21676 from: 1 to: 226

35 ThrProSerValIleThr 41
|||||
171 ACACCATCAGTATTATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK47841

```

```

seq_documentation_block:
ID   AAK47841 standard; DNA; 226 BP.
AC   AAK47841;
XX
XX
DT   06-NOV-2001 (first entry)
XX
XX
XX   Human bone marrow expressed single exon probe SEQ ID NO: 22398.
XX
XX   Human; bone marrow expressed exon; gene expression analysis; probe;
XX   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX   Homo sapiens.
XX
XX   WO200157276-A2.
XX
XX   09-AUG-2001.
XX
XX   30-JAN-2001; 2001WO-US00668.
XX
XX   04-FEB-2000; 2000US-0180312.
XX   26-MAY-2000; 2000US-0207456.
XX   30-JUN-2000; 2000US-0608408.
XX   03-AUG-2000; 2000US-0632366.
XX   21-SEP-2000; 2000US-0234687.
XX   27-SEP-2000; 2000US-0236359.
XX   04-OCT-2000; 2000GB-0024263.
XX
XX   (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX   WPI; 2001-488900/53.
XX
XX   Human genome-derived single exon nucleic acid probes useful for
XX   analyzing gene expression in human bone marrow -
XX
XX   Example 4; SEQ ID NO: 22398; 658pp + Sequence Listing; English.
XX
XX   The present invention provides a number of single exon nucleic acid
XX   probes which are derived from genomic sequences expressed in the human
XX   bone marrow. They can be used to measure gene expression in bone marrow
XX   samples, which may enable the improved diagnosis and treatment of cancers
XX   such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX   the probes of the invention.
XX
XX   Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAK47841 ..
Align seg 1/1 to: AAK47841 from: 1 to: 226
35 ThrProSerValLeuIleThr 41
|||||
171 ACACCATCAGTATTATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI53670
seq_documentation_block:
ID   AAI53670 standard; DNA; 226 BP.
XX
XX
AC   AAI53670;
XX
XX   17-OCT-2001 (first entry)
XX
XX
XX   Probe #22356 used to measure gene expression in human placenta sample.

```

```

XX
XX   Probe; microarray; human; placenta; antenatal diagnosis;
XX   genetic disorder; ss.
XX
XX   Homo sapiens.
XX
XX   WO200157272-A2.
XX
XX   09-AUG-2001.
XX
XX   30-JAN-2001; 2001WO-US00663.
XX
XX   04-FEB-2000; 2000US-0180312.
XX   26-MAY-2000; 2000US-0207456.
XX   30-JUN-2000; 2000US-0608408.
XX   03-AUG-2000; 2000US-0632366.
XX   21-SEP-2000; 2000US-0234687.
XX   27-SEP-2000; 2000US-0236359.
XX   04-OCT-2000; 2000GB-0024263.
XX
XX   (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX   WPI; 2001-48897/53.
XX
XX   Human genome-derived single exon nucleic acid probes useful for
XX   analyzing gene expression in human placenta -
XX
XX   Claim 25; SEQ ID No 22356; 654pp; English.
XX
XX   The present invention relates to single exon nucleic acid probes (SENP).
XX   The present sequence is one such probe. The probes are useful for
XX   producing a microarray for predicting, measuring and displaying gene
XX   expression in samples derived from human placenta. The probes are useful
XX   for antenatal diagnosis of human genetic disorders.
XX
XX   Sequence 226 BP; 71 A; 45 C; 45 G; 65 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAI53670 ..
Align seg 1/1 to: AAI53670 from: 1 to: 226
35 ThrProSerValLeuIleThr 41
|||||
171 ACACCATCAGTATTATAACA 191

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI12381
seq_documentation_block:
ID   AAI12381 standard; cDNA; 262 BP.
XX
XX
AC   AAI12381;
XX
XX   07-DEC-2001 (first entry)
XX
XX   Human breast cancer expressed polynucleotide 4838.
XX
XX   Human; breast cancer; cell marker; cytostatic; ss.
XX
XX   Homo sapiens.
XX
XX   WO200151628-A2.
XX
XX   19-JUL-2001.
XX
XX

```

PF 10-JAN-2001; 2001WO-US00798.
 XX
 PR 14-JAN-2000; 2000US-0176077.
 PR 14-MAR-2000; 2000US-0189167.
 PR 24-MAR-2000; 2000US-0192099.
 PR 29-MAR-2000; 2000US-0193480.
 PR 15-MAY-2000; 2000US-0205230.
 PR 09-JUN-2000; 2000US-0211315.
 PR 25-JUL-2000; 2000US-0220534.
 XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 XX Lillie J, Xu Y, Wang Y, Steinmann K;
 PI WPI; 2001-451856/48.
 DR
 XX
 XX New peptide useful as a marker for the diagnosis of breast cancer -

XX
 XX Claim 1; Page 869; 3695pp; English.
 PS

XX The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 CC

XX Sequence 262 BP; 70 A; 63 C; 64 G; 65 T; 0 other;
 SQ

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAL12381/rev ..
 Align seg 1/1 to reverse of: AAL12381 from: 1 to: 262

159 AsnGlnCysAlaGlyGlyAla 165
 |||||
 76 AACCAAGTGCAGGGGTGCA 56

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAL21257

seq_documentation_block:
 ID AAL21257 standard; cDNA; 262 BP.
 XX
 AC AAL21257;
 XX
 DT 07-DEC-2001 (first entry)
 XX
 DE Human breast cancer expressed polynucleotide 13714.
 XX
 KW Human; breast cancer; cell marker; cytostatic; ss.
 XX
 OS Homo sapiens.
 XX

XX WO200151628-A2.
 XX
 PD 19-JUL-2001.
 XX

PF 10-JAN-2001; 2001WO-US00798.
 XX
 XX 14-JAN-2000; 2000US-0176077.
 PR 14-MAR-2000; 2000US-0189167.
 PR 24-MAR-2000; 2000US-0192099.
 PR 29-MAR-2000; 2000US-0193480.
 PR 15-MAY-2000; 2000US-0205230.
 PR

PR 09-JUN-2000; 2000US-0211315.
 PR 25-JUL-2000; 2000US-0220534.
 XX
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA

XX Lillie J, Xu Y, Wang Y, Steinmann K;
 PI WPI; 2001-451856/48.
 DR
 XX
 XX New peptide useful as a marker for the diagnosis of breast cancer -

XX
 XX Claim 1; Page 2439; 3695pp; English.
 PS

XX The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 CC

XX Sequence 262 BP; 71 A; 63 C; 63 G; 65 T; 0 other;
 SQ

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAL21257/rev ..
 Align seg 1/1 to reverse of: AAL21257 from: 1 to: 262

159 AsnGlnCysAlaGlyGlyAla 165
 |||||
 76 AACCAAGTGCAGGGGTGCA 56

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK65002

seq_documentation_block:
 ID AAK65002 standard; DNA; 264 BP.
 XX
 AC AAK65002;
 XX

DT 06-NOV-2001 (first entry)
 XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19814.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX

OS Homo sapiens.
 XX

XX WO200157182-A2.
 XX

XX 09-AUG-2001.
 XX

PF 17-JAN-2001; 2001WO-US01354.
 XX

XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 15-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR

CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 264 BP; 73 A; 50 C; 46 G; 95 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAK65002 ..

Align seg 1/1 to: AAK65002 from: 1 to: 264

164 GlyAlaAlaMetSerLyshis 170
 |||||
 33 GGAGCAGCAATGCTTAACAT 53

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAV89764

seq_documentation_block:

ID AAV89764 standard; cDNA; 266 BP.

XX
 AC AAV89764;

DT 15-FEB-1999 (first entry)

XX
 DE EST clone CT159.

XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.

XX Homo sapiens.

XX WO9845436-A2.

XX PD 15-OCT-1998.

XX PF 10-APR-1998; 98WO-US06955.

XX PR 10-APR-1997; 97US-0838821.

XX (GEMY) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racie LA, Spaulding V, Treacy M;

XX WPI; 1999-070077/06.

XX New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.

XX Claim 1; Page 322; 618pp; English.

XX The present sequence represents a human expressed sequence tag (EST).
 CC The polynucleotide, which is a secreted EST, and the encoded protein
 CC are predicted to have useful biological activities which would make
 CC them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data is
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating or suppressing activity, haematopoiesis regulating

CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.
 XX
 SQ Sequence 266 BP; 79 A; 38 C; 54 G; 95 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAV89764/rev ..

Align seg 1/1 to reverse of: AAV89764 from: 1 to: 266

12 LeuileSerSerMetLeuVal 18
 |||||
 87 TTAATCTCTTCTATGTTGGTC 67

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:ABA50979

seq_documentation_block:

ID ABA50979 standard; DNA; 282 BP.

XX
 AC ABA50979;

XX 01-FEB-2002 (first entry)

XX Human breast cell single exon nucleic acid probe #9674.

XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/34.

XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -

XX Claim 4; SEQ ID NO 9674; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA50979 ..

Align seg 1/1 to: ABA50979 from: 1 to: 282

100 GlnLeuLeuThrAlaArg 106
 |||||
 24 CAGCTTCTCACCACGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA68959

seq_documentation_block:

ID ABA68959 standard; DNA; 282 BP.

AC ABA68959;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #17264.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

XX Claim 4; SEQ ID NO 17264; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x ABA68959 ..

Align seg 1/1 to: ABA68959 from: 1 to: 282

100 GlnLeuLeuThrAlaArg 106
 |||||
 24 CAGCTTCTCACCACGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA35907

seq_documentation_block:

ID ABA35907 standard; DNA; 282 BP.

XX ABA35907;

XX 23-JAN-2002 (first entry)

XX Probe #14373 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

OS Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

XX Claim 4; SEQ ID NO 14373; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x ABA35907 ..

Align seg 1/1 to: ABA35907 from: 1 to: 282

100 GlnLeuLeuThrThrAlaArg 106
|||||
24 CAGCTTCTCACCACCTGCAAGG 44

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK17286

seq_documentation_block:
ID AAK17286 standard; DNA; 282 BP.

XX AAK17286;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 17277.

XX Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains

XX Example 4; SEQ ID NO: 17277; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is one of the probes of the

XX invention.

XX Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAK17286 ..

Align seg 1/1 to: AAK17286 from: 1 to: 282

100 GlnLeuLeuThrThrAlaArg 106
|||||
24 CAGCTTCTCACCACCTGCAAGG 44

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK43076

seq_documentation_block:

ID AAK43076 standard; DNA; 282 BP.

XX AAK43076;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 17633.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow

XX Example 4; SEQ ID NO: 17633; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples, which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukaemia and myeloma. The present sequence is one of

XX the probes of the invention.

XX Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-674-779-2 x AAK43076 ..

Align seg 1/1 to: AAK43076 from: 1 to: 282

100 GlnLeuLeuThrAlaArg 106
 |||||
 24 CAGCTTCTCACCACGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI23842

seq_documentation_block:

ID AAI23842 standard; DNA; 282 BP.

XX

AC AAI23842;

XX

DT 12-OCT-2001 (first entry)

XX

DE Probe #13775 for gene expression analysis in human cervical cell sample.

XX

KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX

KW cervical cancer; ss.

XX

OS Homo sapiens.

XX

PN WO200157278-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00670.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488901/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

XX

PT analyzing gene expression in human cervical epithelial cells -

XX

PS Claim 25; SEQ ID No 13775; 487pp; English.

XX

CC The present invention relates to human single exon nucleic acid probes

XX

CC (SENP). The present sequence is one such probe. The SENPs are derived

XX

CC from human HeLa cells. The SENPs can be used to produce a single exon

XX

CC microarray, which can be used for measuring human gene expression in a

XX

CC sample derived from human cervical epithelial cells. By measuring gene

XX

CC expression, the probes are therefore useful in grading and/or staging

XX

CC of diseases of the cervix, notably cervical cancer.

XX

CC Note: The sequence data for this patent did not form part of the printed

XX

CC specification, but was obtained in electronic format directly from WIPO

XX

CC at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAI23842 ..

Align seg 1/1 to: AAI23842 from: 1 to: 282

100 GlnLeuLeuThrAlaArg 106

XX

AC

|||||
 24 CAGCTTCTCACCACGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI49154

seq_documentation_block:

ID AAI49154 standard; DNA; 282 BP.

XX

AC AAI49154;

XX

DT 17-OCT-2001 (first entry)

XX

DE Probe #17840 used to measure gene expression in human placenta sample.

XX

KW Probe; microarray; human; placenta; antenatal diagnosis;

XX

KW genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200157272-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00663.

XX

PR 04-FEB-2000; 2000US-0180312.

XX

PR 26-MAY-2000; 2000US-0207456.

XX

PR 30-JUN-2000; 2000US-0608408.

XX

PR 03-AUG-2000; 2000US-0632366.

XX

PR 21-SEP-2000; 2000US-0234687.

XX

PR 27-SEP-2000; 2000US-0236359.

XX

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-48897/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

XX

PT analyzing gene expression in human placenta -

XX

PS Claim 25; SEQ ID No 17840; 654pp; English.

XX

CC The present invention relates to single exon nucleic acid probes (SENP).

XX

CC The present sequence is one such probe. The probes are useful for

XX

CC producing a microarray for predicting, measuring and displaying gene

XX

CC expression in samples derived from human placenta. The probes are useful

XX

CC for antenatal diagnosis of human genetic disorders.

XX

SQ Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAI49154 ..

Align seg 1/1 to: AAI49154 from: 1 to: 282

100 GlnLeuLeuThrAlaArg 106

XX

AC

24 CAGCTTCTCACCACGCAAGG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI09449

seq_documentation_block:

ID AAI09449 standard; DNA; 282 BP.

XX

AC

AAI09449;

XX 09-OCT-2001 (first entry)
 DT Probe #9440 used to measure gene expression in human breast sample.
 DE Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 KW Homo sapiens.
 OS WO200157270-A2.
 PN 09-AUG-2001.
 PD 29-JAN-2001; 2001WO-US00661.
 PF 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 01-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-476286/51.
 XX Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 PT Claim 25; SEQ ID NO 9440; 322pp; English.
 PS The present invention relates to novel single exon nucleic acid probes.
 XX The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 282 BP; 79 A; 70 C; 78 G; 55 T; 0 other;
 SQ

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAI09449 ..
 Align seg 1/1 to: AAI09449 from: 1 to: 282

100 GlnLeuLeuThrThrAlaArg 106
 |||||
 24 CAGCTTCTCACCACCTGCARG 44

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT: AAS61584
 seq_documentation_block:
 ID AAS61584 standard; cDNA; 285 BP.
 XX
 AC AAS61584;
 XX

DT 29-JAN-2002 (first entry)
 XX Lung small cell carcinoma antigen, cDNA #125.
 DE Human; cytostatic; antitumour; lung small cell cancer antigen;
 KW tumour; lung cancer; ss.
 KW Homo sapiens.
 OS WO200177168-A2.
 PN 18-OCT-2001.
 PD 11-APR-2001; 2001WO-US11859.
 PF 11-APR-2000; 2000US-196780P.
 XX 21-JUN-2000; 2000US-213361P.
 PR 01-SEP-2000; 2000US-229763P.
 PR 05-SEP-2000; 2000US-230629P.
 PR 14-SEP-2000; 2000US-232565P.
 PR 19-DEC-2000; 2000US-257037P.
 PR 08-JAN-2001; 2001US-260796P.
 XX (CORI-) CORIXA CORP.
 PA Lodes MJ, Wang T, Mohamath R, Indirias CV;
 XX WPI; 2002-010896/01.
 PI Lung tumour polynucleotide and polypeptides useful in therapy and
 DR diagnosis of cancer especially lung cancer -
 XX Claim 1; Page 166; 295pp; English.
 PS The invention relates to novel isolated lung small cell cancer antigen
 CC polynucleotides (I) and polypeptides (II) used in a method of detecting
 CC cancer in a patient. The method is optionally performed by
 CC utilising oligonucleotides (III), where the biological sample
 CC from the patient is contacted with (III), detecting the amount of
 CC polynucleotide hybridised to (III) in the sample and comparing the
 CC amount of polynucleotide to a predetermined cut-off value and thereby
 CC determining cancer in a patient. (I), (II) or antigen-presenting cells
 CC expressing (II) is useful for stimulating and/or expanding T cells
 CC specific for a tumour protein. The method comprises contacting T cells
 CC with one of the components under conditions to permit the stimulation
 CC and/or expansion of the cells. A composition comprising (I) is useful for
 CC stimulating an immune response in a patient and for inhibiting the
 CC development of a cancer especially lung cancer in a patient. An
 CC isolated T cell population is useful for removing tumour cells from the
 CC biological sample and for inhibiting the development of cancer in a
 CC patient. AAS61460-AAS61874 represent novel human lung small cell
 CC cancer antigen coding sequences of the invention.
 XX Sequence 285 BP; 81 A; 59 C; 55 G; 77 T; 13 other;
 SQ

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAS61584 ..
 Align seg 1/1 to: AAS61584 from: 1 to: 285

127 IleValProThrLeuHisLeu 133
 |||||
 98 ATAGTACCACCCCTCATCTC 118

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT: ABA11586
 seq_documentation_block:

ID XX ABAll586 standard; cDNA; 295 BP.
AC XX ABAll586;
XX XX
XX XX 23-JAN-2002 (first entry)
DT DT
DE DE Human nervous system related polynucleotide SEQ ID NO 593.
XX XX
XX XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
KW KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW KW antiparkinsonian; antiskinking; antianaemic; antiarthritic; cancer;
KW KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
OS OS
XX XX Homo sapiens.
PN PN WO200159063-A2.
XX XX
PD PD 16-AUG-2001.
XX XX
PF PF 17-JAN-2001; 2001WO-US01334.
XX XX
PR PR 31-JAN-2000; 2000US-0179065.
PR PR 04-FEB-2000; 2000US-0180628.
PR PR 24-FEB-2000; 2000US-0184664.
PR PR 02-MAR-2000; 2000US-0186350.
PR PR 16-MAR-2000; 2000US-0189874.
PR PR 17-MAR-2000; 2000US-0190076.
PR PR 18-APR-2000; 2000US-0198123.
PR PR 19-MAY-2000; 2000US-0205515.
PR PR 07-JUN-2000; 2000US-0209467.
PR PR 28-JUN-2000; 2000US-0214886.
PR PR 30-JUN-2000; 2000US-0215135.
PR PR 07-JUL-2000; 2000US-0216647.
PR PR 07-JUL-2000; 2000US-0216880.
PR PR 11-JUL-2000; 2000US-0217487.
PR PR 11-JUL-2000; 2000US-0217496.
PR PR 14-JUL-2000; 2000US-0218290.
PR PR 26-JUL-2000; 2000US-0220963.
PR PR 26-JUL-2000; 2000US-0220964.
PR PR 14-AUG-2000; 2000US-0224518.
PR PR 14-AUG-2000; 2000US-0224519.
PR PR 14-AUG-2000; 2000US-0225213.
PR PR 14-AUG-2000; 2000US-0225214.
PR PR 14-AUG-2000; 2000US-0225266.
PR PR 14-AUG-2000; 2000US-0225267.
PR PR 14-AUG-2000; 2000US-0225268.
PR PR 14-AUG-2000; 2000US-0225270.
PR PR 14-AUG-2000; 2000US-0225447.
PR PR 14-AUG-2000; 2000US-0225757.
PR PR 14-AUG-2000; 2000US-0225758.
PR PR 14-AUG-2000; 2000US-0225759.
PR PR 18-AUG-2000; 2000US-0226279.
PR PR 22-AUG-2000; 2000US-0226681.
PR PR 22-AUG-2000; 2000US-0226686.
PR PR 22-AUG-2000; 2000US-0227182.
PR PR 23-AUG-2000; 2000US-0227009.
PR PR 30-AUG-2000; 2000US-0228924.
PR PR 01-SEP-2000; 2000US-0229287.
PR PR 01-SEP-2000; 2000US-0229343.
PR PR 01-SEP-2000; 2000US-0229344.
PR PR 01-SEP-2000; 2000US-0229345.
PR PR 05-SEP-2000; 2000US-0229509.
PR PR 05-SEP-2000; 2000US-0229513.
PR PR 06-SEP-2000; 2000US-0230437.
PR PR 06-SEP-2000; 2000US-0230438.
PR PR 08-SEP-2000; 2000US-0231242.
PR PR 08-SEP-2000; 2000US-0231243.
PR PR 08-SEP-2000; 2000US-0231244.
PR PR 08-SEP-2000; 2000US-0231413.
PR PR 08-SEP-2000; 2000US-0231414.
PR PR 08-SEP-2000; 2000US-0232080.
PR PR 08-SEP-2000; 2000US-0232081.
PR PR 12-SEP-2000; 2000US-0231968.
PR PR 14-SEP-2000; 2000US-0232397.
PR PR 14-SEP-2000; 2000US-0232398.
PR PR 14-SEP-2000; 2000US-0232399.
PR PR 14-SEP-2000; 2000US-0232400.
PR PR 14-SEP-2000; 2000US-0232401.
PR PR 14-SEP-2000; 2000US-0233063.
PR PR 14-SEP-2000; 2000US-0233064.
PR PR 14-SEP-2000; 2000US-0233065.
PR PR 21-SEP-2000; 2000US-0234223.
PR PR 21-SEP-2000; 2000US-0234274.
PR PR 25-SEP-2000; 2000US-0234997.
PR PR 25-SEP-2000; 2000US-0234998.
PR PR 26-SEP-2000; 2000US-0235484.
PR PR 27-SEP-2000; 2000US-0235834.
PR PR 27-SEP-2000; 2000US-0235836.
PR PR 29-SEP-2000; 2000US-0236327.
PR PR 29-SEP-2000; 2000US-0236367.
PR PR 29-SEP-2000; 2000US-0236368.
PR PR 29-SEP-2000; 2000US-0236369.
PR PR 29-SEP-2000; 2000US-0236370.
PR PR 02-OCT-2000; 2000US-0236802.
PR PR 02-OCT-2000; 2000US-0237037.
PR PR 02-OCT-2000; 2000US-0237038.
PR PR 02-OCT-2000; 2000US-0237039.
PR PR 02-OCT-2000; 2000US-0237040.
PR PR 13-OCT-2000; 2000US-0239935.
PR PR 13-OCT-2000; 2000US-0239937.
PR PR 20-OCT-2000; 2000US-0240960.
PR PR 20-OCT-2000; 2000US-0241785.
PR PR 20-OCT-2000; 2000US-0241786.
PR PR 20-OCT-2000; 2000US-0241787.
PR PR 20-OCT-2000; 2000US-0241808.
PR PR 20-OCT-2000; 2000US-0241809.
PR PR 20-OCT-2000; 2000US-0241826.
PR PR 20-OCT-2000; 2000US-0242221.
PR PR 01-NOV-2000; 2000US-0244617.
PR PR 08-NOV-2000; 2000US-0244674.
PR PR 08-NOV-2000; 2000US-0246475.
PR PR 08-NOV-2000; 2000US-0246476.
PR PR 08-NOV-2000; 2000US-0246477.
PR PR 08-NOV-2000; 2000US-0246478.
PR PR 08-NOV-2000; 2000US-0246523.
PR PR 08-NOV-2000; 2000US-0246524.
PR PR 08-NOV-2000; 2000US-0246525.
PR PR 08-NOV-2000; 2000US-0246526.
PR PR 08-NOV-2000; 2000US-0246527.
PR PR 08-NOV-2000; 2000US-0246528.
PR PR 08-NOV-2000; 2000US-0246532.
PR PR 08-NOV-2000; 2000US-0246533.
PR PR 08-NOV-2000; 2000US-0246610.
PR PR 08-NOV-2000; 2000US-0246611.
PR PR 17-NOV-2000; 2000US-0246613.
PR PR 17-NOV-2000; 2000US-0249207.
PR PR 17-NOV-2000; 2000US-0249208.
PR PR 17-NOV-2000; 2000US-0249209.
PR PR 17-NOV-2000; 2000US-0249210.
PR PR 17-NOV-2000; 2000US-0249211.
PR PR 17-NOV-2000; 2000US-0249212.
PR PR 17-NOV-2000; 2000US-0249213.
PR PR 17-NOV-2000; 2000US-0249214.
PR PR 17-NOV-2000; 2000US-0249215.
PR PR 17-NOV-2000; 2000US-0249216.
PR PR 17-NOV-2000; 2000US-0249217.
PR PR 17-NOV-2000; 2000US-0249218.
PR PR 17-NOV-2000; 2000US-0249244.
PR PR 17-NOV-2000; 2000US-0249245.
PR PR 17-NOV-2000; 2000US-0249264.
PR PR 17-NOV-2000; 2000US-0249265.
PR PR 17-NOV-2000; 2000US-0249297.
PR PR 17-NOV-2000; 2000US-0249299.
PR PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX WPI; 2001-541565/60.
 DR P-PSDB; ABB15260.
 XX
 DR Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 PT
 XX
 PS Claim 1; SEQ ID NO 593; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AB11004-ABA21534) and proteins
 CC (AB114678-ABB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 295 BP; 94 A; 31 C; 54 G; 114 T; 2 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x ABA11586/rev ..
 Align seg 1/1 to reverse of: ABA11586 from: 1 to: 295

34 LysThrProSerValLeulle 40
 |||||
 31 AAACACCGCTGTATTAATT 11

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV37035

seq_documentation_block:
 ID AAV37035 standard; DNA; 312 BP.
 XX
 AC AAV37035;
 XX
 DT 04-SEP-1998 (first entry)
 XX

DE Streptococcus salivarius sequence used to design primers and probes.
 XX
 KW Detection; bacterial antibiotic resistance gene; bacteria;
 KW fungal species; identification; ds.
 XX
 OS Streptococcus salivarius.
 XX
 PN WO9820157-A2.
 XX
 PD 14-MAY-1998.
 XX
 XX 04-NOV-1997; 97WO-CA00829.
 PF
 XX 04-NOV-1996; 96US-0743637.
 PR
 XX (IDI-) IDI INFECTIO DIAGNOSTIC INC.
 PA
 XX Bergeron MG, Ouellette M, Picard FJ, Roy PH;
 PI
 XX WPI; 1998-286967/25.
 DR
 XX
 PT Use of oligo:nucleotide primers and probes - for detection,
 PT identification and quantification of bacteria, fungi and bacterial
 PT antibiotic resistance gene(s)
 XX
 XX Claim 10; Page 87; 167pp; English.
 PS
 CC The present sequence was used to design primers and probes which
 CC are used in the course of the invention. The specification describes the
 CC use of probes and/or amplification primers which are specific, ubiquitous
 CC and sensitive for determining the presence and amount of nucleic acids
 CC from a bacterial antibiotic resistance gene and specific bacterial and
 CC fungal species in any sample suspected of containing the bacterial or
 CC fungal nucleic acids, where each of the nucleic acid or variant or part
 CC comprises a selected target region hybridisable with the probes or
 CC primers. The method of use comprises contacting the sample with the
 CC probes or primers and detecting the presence of hybridised probes or
 CC amplified products as an indication of the presence of the specific
 CC bacterial or fungal species and bacterial antibiotic resistance genes.
 CC The methods and products can be used to detect and identify the bacterial
 CC and fungal species and genera and determine the bacterial resistance to
 CC antibiotics.
 XX
 SQ Sequence 312 BP; 81 A; 54 C; 80 G; 97 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-674-779-2 x AAV37035 ..
 Align seg 1/1 to: AAV37035 from: 1 to: 312

58 SerHisValGlyLeuGlnAla 64
 |||||
 181 AGTCATGTAGGACTTCAAGCG 201

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH01713

seq_documentation_block:
 ID AAH01713 standard; DNA; 312 BP.
 XX
 AC AAH01713;
 XX
 DT 24-JUL-2001 (first entry)
 XX
 XX Streptococcus salivarius nucleotide sequence SEQ ID NO:1706.
 DE
 XX Species specific; genus specific; family specific; probe; detection;
 KW identification; algal; archaeal; bacterial; fungal; parasitica;
 KW

KW microorganism; diagnosis; translation elongation factor Tu; toxin;
 KW translation elongation factor G; RecA recombinase; resistance;
 KW catalytic subunit of proton-translocating ATPase; antimicrobial;
 KW vaccine; primer; ds.
 XX
 XX Streptococcus salivarius.
 XX
 XX WO200123604-A2.
 XX
 XX 05-APR-2001.
 XX
 XX 28-SEP-2000; 2000WO-CA01150.
 XX
 XX 28-SEP-1999; 99CA-2283458.
 PR 19-MAY-2000; 2000CA-2307010.
 XX
 XX (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
 XX
 XX Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
 PI Picard FJ, Roy PH;
 XX
 XX WPI; 2001-245006/25.
 XX
 XX Nucleic acid sequences are used to generate universal probes and
 PT primers which can be used to identify and detect the presence of algal,
 PT archaeal, bacterial, fungal and parasitological species in a test sample -
 XX
 XX Disclosure; Page 1288; 1580pp; English.
 XX
 CC The present invention describes a method for generating a repository of
 CC nucleic acids of tuf, fus, atpD and/or recA genes from which probes
 CC and/or primers are derived. The method comprises amplifying the nucleic
 CC acids of determined algal, archaeal, bacterial, fungal and parasitological
 CC species with a combination of defined primer pairs. The method can be
 CC used for producing probes and/or primers for detecting one or more
 CC related microorganisms e.g. algae, archaea, bacteria, fungi and
 CC parasites, for universal detection and for specific and ubiquitous
 CC detection and identification of an algal, archaeal, bacterial, fungal
 CC and parasitological species, genus, family and group. A nucleic acid (I)
 CC obtained using the method of the invention can be used for the universal
 CC detection of any bacterium, fungus or parasite in a sample and for the
 CC detection of at least one antimicrobial agent resistance gene or at
 CC least one toxin gene. Hexa nucleic acids are used for the specific and
 CC ubiquitous detection and for identification of Streptococcus pneumoniae.
 CC (I) can be used to design a therapeutic agent which is effective against
 CC microorganisms. Microbial species or genus or family or phylum or group
 CC which can be detected include Abiotrophia adiacens, Bordetella sp.,
 CC Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
 CC Mycobacteriaceae family, Pseudomonads group, Streptococcus sp.,
 CC Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
 CC provides faster results than substrate specificity tests as results can
 CC be determined in an hour and improved accuracy is also achieved.
 CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
 CC which are given in the exemplification of the present invention.
 XX
 SQ Sequence 312 BP; 81 A; 54 C; 80 G; 97 T; 0 other;

alignment_scores:
 Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAH01713 ..

Align seg 1/1 to: AAH01713 from: 1 to: 312

58 SerHisValGlyLeuGlnAla 64
 |||||
 181 AGTCATGTAGGACTTCAAGCG 201

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:AAH16270

seq_documentation_block:
 ID AAT16270 standard; DNA; 318 BP.

XX AAT16270;

XX 01-AUG-1996 (first entry)

XX 0.3 kb recA fragment from pGEM-T.

XX autonomous replication; inverted repeat; insertion sequence;
 KW open reading frame; plasmid; isolation; amino acid biosynthesis;
 KW marker gene; ds.

XX Brevibacterium flavus strain MJ-233.

XX JP07327680-A.

XX 19-DEC-1995.

XX 07-JUN-1994; 94JP-0124852.

XX 07-JUN-1994; 94JP-0124852.

XX (MITU) MITSUBISHI CHEM CORP.

XX WPI; 1996-072338/08.

XX P-PSDB; R9251.

XX Plasmid which cannot be replicated autonomously in Coryneform
 PT bacteria - useful for isolating genes involved in amino acid
 PT biosynthesis

XX Disclosure; Page 15; 16pp; Japanese.

XX A plasmid which cannot be replicated autonomously in a Coryneform
 CC bacterium carries a DNA region having at least one inverted repeat
 CC (IR) present upstream or downstream of an open reading frame within
 CC an inserted sequence. The IRs are derived from the 5' upstream and
 CC 3' downstream regions of a marker gene derived from a Coryneform
 CC bacterium. The plasmid is used in a method for obtaining a mutant in
 CC which a Coryneform bacterium is transformed with the plasmid and the
 CC strain expressing the marker gene is separated by using the marker gene
 CC of the transformant as the index. Genes relating to amino acid
 CC biosynthesis can be easily isolated by using the mutant. The present
 CC sequence is a 0.3 kb PCR fragment contg. the recA gene from
 CC Brevibacterium flavus strain MJ-233. A recA gene-deleted mutant strain
 CC was prepd. for use in the method of the invention.

XX Sequence 318 BP; 72 A; 82 C; 91 G; 73 T; 0 other;

alignment_scores:

Quality: 7.00 Length: 7
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-674-779-2 x AAT16270 ..

Align seg 1/1 to: AAT16270 from: 1 to: 318

58 SerHisValGlyLeuGlnAla 64
 |||||
 187 AGCCACGTTGGCTTCAGGCC 207

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH01012

seq_documentation_block:

ID AAH01012 standard; DNA; 332 BP.

XX AAH01012;

XX

```

DT 24-JUL-2001 (first entry)
XX
DE Bacillus cereus nucleotide sequence SEQ ID NO:1003.
XX
KW Species specific; genus specific; family specific; probe; detection;
KW identification; algal; archaeal; bacterial; fungal; parasitological;
KW microorganism; diagnosis; translation elongation factor Tu; toxin;
KW translation elongation factor G; RecA recombinase; resistance;
KW catalytic subunit of proton-translocating ATPase; antimicrobial;
KW vaccine; primer; ds.
XX
OS Bacillus cereus.
XX
PN WO200123604-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-CA01150.
XX
PR 28-SEP-1999; 99CA-2283458.
XX
PR 19-MAY-2000; 2000CA-2307010.
XX
PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
PI Picard FJ, Roy PH;
XX
XX WPI; 2001-245006/25.
XX
Nucleic acid sequences are used to generate universal probes and
primers which can be used to identify and detect the presence of algal,
archaeal, bacterial, fungal and parasitological species in a test sample -
Claim 26; Page 960; 1580pp; English.
XX
The present invention describes a method for generating a repository of
nucleic acids of tuf, fus, atpD and/or recA genes from which probes
and/or primers are derived. The method comprises amplifying the nucleic
acids of determined algal, archaeal, bacterial, fungal and parasitological
species with a combination of defined primer pairs. The method can be
used for producing probes and/or primers for detecting one or more
related microorganisms e.g. algae, archaea, bacteria, fungi and
parasites, for universal detection and for specific and ubiquitous
detection and identification of an algal, archaeal, bacterial, fungal
and parasitological species, genus, family and group. A nucleic acid (I)
obtained using the method of the invention can be used for the universal
detection of any bacterium, fungus or parasite in a sample and for the
detection of at least one antimicrobial agent resistance gene or at
least one toxin gene. hexA nucleic acids are used for the specific and
ubiquitous detection and for identification of Streptococcus pneumoniae.
(I) can be used to design a therapeutic agent which is effective against
microorganisms. Microbial species or genus or family or phylum or group
which can be detected include Abiotrophia adiacens, Bordetella sp.,
Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
Mycobacterium sp., Pseudomonas group, Streptococcus sp.,
Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
provides faster results than substrate specificity tests as results can
be determined in an hour and improved accuracy is also achieved.
CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC which are given in the exemplification of the present invention.
XX
Sequence 332 BP; 115 A; 60 C; 74 G; 83 T; 0 other;
XX
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
XX
alignment_block:
US-09-674-779-2 x AAH01012
XX
Align seg 1/1 to: AAH01012 from: 1 to: 332

```

```

58 SerHisValGlyLeuGlnAla 64
|||||
191 TCACACGTAGGTTTACAAGCA 211

```

```
seq_name: /STDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAH01298
```

```
seq_documentation_block:
```

```
ID AAH01298 standard; DNA; 336 BP.
```

```
XX AC AAH01298;
```

```
XX XX 24-JUL-2001 (first entry)
```

```
XX Bacillus thuringiensis nucleotide sequence SEQ ID NO:1289.
```

```

XX Species specific; genus specific; family specific; probe; detection;
KW identification; algal; archaeal; bacterial; fungal; parasitological;
KW microorganism; diagnosis; translation elongation factor Tu; toxin;
KW translation elongation factor G; RecA recombinase; resistance;
KW catalytic subunit of proton-translocating ATPase; antimicrobial;
KW vaccine; primer; ds.
XX
XX Bacillus thuringiensis.
XX
XX WO200123604-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-CA01150.
XX
XX 28-SEP-1999; 99CA-2283458.
XX
XX 19-MAY-2000; 2000CA-2307010.
XX
XX (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
XX
XX Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
XX Picard FJ, Roy PH;
XX
XX WPI; 2001-245006/25.
XX
Nucleic acid sequences are used to generate universal probes and
primers which can be used to identify and detect the presence of algal,
archaeal, bacterial, fungal and parasitological species in a test sample -
Claim 26; Page 1105; 1580pp; English.
XX
The present invention describes a method for generating a repository of
nucleic acids of tuf, fus, atpD and/or recA genes from which probes
and/or primers are derived. The method comprises amplifying the nucleic
acids of determined algal, archaeal, bacterial, fungal and parasitological
species with a combination of defined primer pairs. The method can be
used for producing probes and/or primers for detecting one or more
related microorganisms e.g. algae, archaea, bacteria, fungi and
parasites, for universal detection and for specific and ubiquitous
detection and identification of an algal, archaeal, bacterial, fungal
and parasitological species, genus, family and group. A nucleic acid (I)
obtained using the method of the invention can be used for the universal
detection of any bacterium, fungus or parasite in a sample and for the
detection of at least one antimicrobial agent resistance gene or at
least one toxin gene. hexA nucleic acids are used for the specific and
ubiquitous detection and for identification of Streptococcus pneumoniae.
(I) can be used to design a therapeutic agent which is effective against
microorganisms. Microbial species or genus or family or phylum or group
which can be detected include Abiotrophia adiacens, Bordetella sp.,
Corynebacterium sp., Enterobacteriaceae group, Escherichia coli,
Mycobacterium sp., Pseudomonas group, Streptococcus sp.,
Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests
provides faster results than substrate specificity tests as results can
be determined in an hour and improved accuracy is also achieved.
CC AAH00010 to AAH002304 represent nucleotide sequences and primers/probes
CC which are given in the exemplification of the present invention.
XX

```


SQ Sequence 336 BP; 111 A; 61 C; 75 G; 85 T; 4 other;

alignment_scores:

Quality:	7.00	Length:	7
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-674-779-2 x AAH01298 ..

Align seg 1/1 to: AAH01298 from: 1 to: 336

58 SerHisValGlyLeuGlnAla 64

|||||

191 TCACACGTAGTTTACAGCA 211

